BIOSIGNALS 2008

International Conference on
Bio-inspired Systems and Signal Processing

Proceedings

Volume I

Funchal, Madeira - Portugal · 28 - 31 January, 2008

TECHNICAL CO-SPONSORSHIP BY

IEEE EMB SIGART

CO-ORGANIZED BY

INSTICC UNIVERSIDADE da MADEIRA

IN COOPERATION WITH

aaai

INSTICC PRESS

BIOSIGNALS 2008

Proceedings of the
First International Conference on
Bio-inspired Systems and Signal Processing

Volume 1

Funchal, Madeira - Portugal

January 28 – 31, 2008

Co-organized by
INSTICC – Institute for Systems and Technologies of Information, Control and Communication
and
UMA – University of Madeira

Technical Co-sponsorship by
IEEE EMB – Engineering in Medicine and Biology Society
and
ACM SIGART- Special Interest Group on Artificial Intelligence

In Cooperation with
AAAI – Association for the Advancement of Artificial Intelligence

Edited by Pedro Encarnação and António Veloso

Printed in Portugal

ISBN: 978-989-8111-18-0

Depósito Legal: 268555/07

http://www.biosignals.org

secretariat@biosignals.org

BIOSIGNALS is part of BIOSTEC – International Joint Conference on Biomedical Engineering Systems and Technologies

BRIEF CONTENTS

INVITED SPEAKERS

Sérgio Cerutti

Polytechnic University of Milan

Italy

Kevin Warwick

University of Reading

U.K.

Fernando Henrique Lopes da Silva

University of Amsterdam

The Netherlands

Vipul Kashyap

Partners HealthCare System, Clinical Informatics R&D

U.S.A.

David Hall

Research Triangle Institute in North Carolina

U.S.A.

Albert Cook

University of Alberta, Faculty of Rehabilitation Medicine

Canada

ORGANIZING AND STEERING COMMITTEES

CONFERENCE CO-CHAIRS

Ana Fred, IST- Technical University of Lisbon, Portugal

Joaquim Filipe, INSTICC / Polytechnic Institute of Setúbal, Portugal

Hugo Gamboa, Telecommunications Institute, Portugal

Jorge Cardoso, University of Madeira - UMA / Madeira, Portugal

PROGRAM CO-CHAIRS

Pedro Encarnação, Catholic Portuguese University, Portugal

António Veloso, FMH, Universidade Técnica de Lisboa, Portugal

LOCAL CHAIR

Paulo Sampaio, University of Madeira - UMA, Portugal

PROCEEDINGS PRODUCTION

Paulo Brito, INSTICC, Portugal

Marina Carvalho, INSTICC, Portugal

Helder Coelhas, INSTICC, Portugal

Vera Coelho, INSTICC, Portugal

Andreia Costa, INSTICC, Portugal

Bruno Encarnação, INSTICC, Portugal

Bárbara Lima, INSTICC, Portugal

Vitor Pedrosa, INSTICC, Portugal

Vera Rosário, INSTICC, Portugal

Mónica Saramago, INSTICC, Portugal

CD-ROM PRODUCTION

Paulo Brito, INSTICC, Portugal

GRAPHICS PRODUCTION AND WEB DESIGNER

Marina Carvalho, INSTICC, Portugal

SECRETARIAT AND WEBMASTER

Marina Carvalho, INSTICC, Portugal

PROGRAM COMMITTEE

Andrew Adamatzky, University of the West of England, Bristol, U.K.

Cedric Archambeau, University College London, U.K.

Magdy Bayoumi, University of Louisiana, U.S.A.

Peter Bentley, University College London, U.K.

Paolo Bonato, Harvard Medical School, U.S.A.

Marleen de Bruijne, University of Copenhagen, Denmark

Zehra Cataltepe, Istanbul Technical University, Turkey

Gert Cauwenberghs, University of California San Diego, U.S.A.

Mujdat Cetin, Sabanci University, Turkey

Wael El-Deredy, University of Manchester, U.K.

Eran Edirisinghe, Loughborough University, U.K.

Eugene Fink, Carnegie Mellon University, U.S.A.

Luc Florack, Eindhoven University of Technology, the Netherlands

David Fogel, Natural Selection, Inc., U.S.A.

Alejandro Frangi, Universitat Pompeu Fabra, Spain

Sebastià Galmés, Universitat de les Illes Balears, Spain

Aaron Golden, National University of Ireland, Galway, Ireland

Rodrigo Guido, University of São Paulo, Brazil

Bin He, University of Minnesota, U.S.A.

Roman Hovorka, University of Cambridge, U.K.

Helmut Hutten, Graz University of Technology, Austria

Christopher James, University of Southampton, U.K.

Lars Kaderali, University of Heidelberg, Germany

Gunnar W. Klau, Free University Berlin, Germany

Alex Kochetov, Institute of Cytology and Genetics, Russian Federation

T. Laszlo Koczy, Budapest University of Technology and Economics, Hungary

Georgios Kontaxakis, Universidad Politécnica de Madrid, Spain

Igor Kotenko, St. Petersburg Institute for Informatics and Automation, Russian Federation

Narayanan Krishnamurthi, Arizona State University, U.S.A.

Arjan Kuijper, RICAM, Austria

Andrew Laine, Columbia University, U.S.A.

Anna T. Lawniczak, University of Guelph, Canada

Jason J. S. Lee, National Yang-Ming University, Taiwan, Province of China

Kenji Leibnitz, Osaka University, Japan

Marco Loog, University of Copenhagen, Denmark

David Lowe, Aston University, U.K.

Mahdi Mahfouf, The University of Sheffield, U.K.

Luigi Mancini, Universita di Roma La Sapienza, Italy

Elena Marchiori, VUA, the Netherlands

Fabio Martinelli, IIT-CNR, Italy

Martin Middendorf, University of Leipzig, Germany

Mariofanna Milanova, University of Arkansas at Little Rock, U.S.A.

Charles Mistretta, University of Wisconsin, U.S.A.

Gabor Mocz, University of Hawaii, U.S.A.

Kayvan Najarian, University of North Carolina at Charlotte, U.S.A.

Tadashi Nakano, University of California, Irvine, U.S.A.

Asoke K. Nandi, The University of Liverpool, U.K.

Antti Niemistö, Institute for Systems Biology, U.S.A.

Maciej J. Ogorzalek, Jagiellonian University, Poland

Kazuhiro Oiwa, National Institute of Information and Communications Technology, Japan

Jean-Chistophe Olivo-Marin, Institut Pasteur, France

Ernesto Pereda, University of La Laguna, Spain

Leif Peterson, The Methodist Hospital, U.S.A.

Gert Pfurtscheller, Graz University of Technology, Austria

Vitor Fernão Pires, ESTSetubal / IPS, Portugal

Chi-Sang Poon, MIT, U.S.A.

José Principe, University of Florida, U.S.A.

PROGRAM COMMITTEE (CONT.)

Chi-Sang Poon, MIT, U.S.A.

Nikolaus Rajewsky, Max Delbruck Center for Molecular Medicine, Germany

Dick de Ridder, Delft University of Technology, the Netherlands

Joel Rodrigues, Institue of Telecommunications / University of Beira Interior, Portugal

Marcos Rodrigues, Sheffield Hallam University, U.K.

Virginie Ruiz, University of Reading, U.K.

Heather Ruskin, Dublin City University, Ireland

William Zev Rymer, RIC/Northwestern University, U.S.A.

Gerald Schaefer, Aston University, U.K.

Dragutin Sevic, Institute of Physics, Belgrade, Serbia

Iryna Skrypnyk, University of Jyvaskyla, Finland

Alan A. Stocker, New York University, U.S.A.

Jun Suzuki, University of Massachusetts, Boston, U.S.A.

Andrzej Swierniak, Silesian University of Technology, Poland

Boleslaw Szymanski, RPI, U.S.A.

Asser Tantawi, IBM, U.S.A.

Lionel Tarassenko, University of Oxford, U.K.

Gianluca Tempesti, University of York, U.K.

Anna Tonazzini, Consiglio Nazionale delle Ricerche - CNR, Italy

Duygu Tosun, Laboratory of Neuro Imaging, U.S.A.

Bart Vanrumste, Katholieke Hogeschool Kempen, Belgium

Didier Wolf, CRAN CNRS UMR 7039, France

Andrew Wood, Swinburne University of Technology, Australia

Guang-Zhong Yang, Imperial College London, U.K.

Eckart Zitzler, ETH Zurich, Switzerland

AUXILIARY REVIEWERS

Qi Duan, Columbia University, U.S.A.

Soo-Yeon Ji, Virginia Commonwealth University, U.S.A.

Yuri Orlov, Genome Institute of Singapore, Singapore

Ting Song, Columbia University, U.S.A.

Bruno N. Di Stefano, Nuptek Systems Ltd., Canada

SELECTED PAPERS BOOK

A number of selected papers presented at BIOSIGNALS 2008 will be published by Springer, in a book entitled Biomedical Engineering Systems and Technologies. This selection will be done by the conference co-chairs and program co-chairs, among the papers actually presented at the conference, based on a rigorous review by the BIOSTEC 2008 program committee members.

OFFICIAL CARRIER

FOREWORD

This volume contains the proceedings of the *First International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS 2008), organized by the Institute for Systems and Technologies of Information Control and Communication (INSTICC) and the University of Madeira, technically co-sponsored by the IEEE Engineering in Medicine and Biology Society (EMB) and in cooperation with AAAI.

The purpose of the *International Conference on Bio-inspired Systems and Signal Processing* is to bring together researchers and practitioners from multiple areas of knowledge, including biology, medicine, engineering and other physical sciences, interested in studying and using models and techniques inspired from or applied to biological systems. A diversity of signal types can be found in this area, including image, audio and other biological sources of information. The analysis and use of these signals is a multidisciplinary area including signal processing, pattern recognition and computational intelligence techniques, amongst others.

BIOSIGNALS is one of three integrated conferences that are co-located and constitute the International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC). The other two component conferences are HEALTHINF (International Conference on Health Informatics) and BIODEVICES (International Conference on Biomedical Electronics and Devices).

The joint conference, BIOSTEC, has received 494 paper submissions from more than 40 countries in all continents. 65 papers were published and presented as full papers, i.e. completed work (8 pages/30' oral presentation), 189 papers reflecting work-in-progress or position papers were accepted for short presentation, and another 86 contributions were accepted for poster presentation. These numbers, leading to a "full-paper" acceptance ratio below 14% and a total oral paper presentations acceptance ratio below 52%, show the intention of preserving a high quality forum for the next editions of this conference.

The conference included a panel and six invited talks delivered by internationally distinguished speakers, namely: Sergio Cerutti, Kevin Warwick, F. H. Lopes da Silva, Vipul Kashyap, David Hall and Albert Cook. Their participation has positively contributed to reinforce the overall quality of the Conference and to provide a deeper understanding of the field of Biomedical Engineering Systems and Technologies.

The proceedings of the conference will be indexed by several major indices including DBLP, INSPEC and ISI-Proceedings and it will also be submitted for indexing to EI. A book with the revised versions of a short list of selected papers from the conference will be published by Springer-Verlag in the new CS book series: Communications in Computer and Information Science (CCIS). Additionally, a special issue of the IEEE Transactions on Biomedical Circuits and Systems will be edited based on the very best papers of the conference.

FOREWORD (CONT.)

The program for this conference required the dedicated effort of many people. Firstly, we must thank the authors, whose research and development efforts are recorded here. Secondly, we thank the members of the program committee and the additional reviewers for their diligence and expert reviewing. Thirdly, we thank the keynote speakers for their invaluable contribution and for taking the time to synthesise and prepare their talks. Fourthly, we thank the program chairs, Pedro Encarnação and António Veloso, whose collaboration was much appreciated. Finally, special thanks to all the members of the INSTICC team, especially Marina Carvalho at the conference secretariat, and the local organising committee from the University of Madeira, especially Jorge Cardoso and Paulo Sampaio, whose collaboration was fundamental for the success of this conference.

This year, the organization will distribute two paper awards at the conference closing session: the best paper award and the best student paper award. The decision was mainly based on the paper classifications provided by the Program Committee.

We wish you all an exciting conference and an unforgettable stay in the lovely island of Madeira. We hope to meet you again next year for the 2nd BIOSIGNALS, details of which are available at http://www.biosignals.org.

Joaquim Filipe

INSTICC/Polytechnic Institute of Setúbal

CONTENTS

POSTERS

INVITED SPEAKERS

KEYNOTE LECTURES

MULTIVARIATE, MULTIORGAN AND MULTISCALE INTEGRATION OF INFORMATION IN BIOMEDICAL SIGNAL PROCESSING

Sergio Cerutti
Department of Bioengineering, Polytechnic University, Milano, Italy
sergio.cerutti@polimi.it

Abstract: Biomedical signals carry important information about the behavior of the living systems under studying. A proper processing of these signals allows in many instances to obtain useful physiological and clinical information. Many advanced algorithms of signal and image processing have recently been introduced in such an advanced area of research and therefore important selective information is obtainable even in presence of strong sources of noise or low signal/noise ratio. Traditional stationary signal analysis together with innovative methods of investigation of dynamical properties of biological systems and signals in second-order or in higher-order approaches (i.e., in time-frequency, time-variant and time-scale analysis, as well as in non linear dynamics analysis) provide a wide variety of even complex processing tools for information enhancement procedures. Another important innovative aspect is also remarked: the integration between signal processing and modeling of the relevant biological systems is capable to directly attribute patho-physiological meaning to the parameters obtained from the processing and viceversa the modeling fitting could certainly be improved by taking into account the results from signal processing procedure. Such an integration process could comprehend parameters and observations detected at different scales, at different organs and with different modalities. This approach is reputed promising for obtaining an olistic view of the patient rather than an atomistic one which considers the whole as a simple sum of the single component parts.

BRIEF BIOGRAPHY

Sergio Cerutti is Professor in Biomedical Signal and Data Processing at the Department of Bioengineering of the Polytechnic University in Milano, Italy. In the period 2000-2006 he has been the Chairman of the same Department. His research interests are mainly in the following topics: biomedical signal processing (ECG, blood pressure signal and respiration, cardiovascular variability signals, EEG and evoked potentials), neurosciences and cardiovascular modelling. In his research activity he has put emphasis on the integration of information at different modalities, at different sources and at different scales in various physiological systems. Since 1983 he has taught a course at a graduate and a doc level on Biomedical Signal Processing and Modelling at Engineering Faculties (Milano and Roma) as well as at Specialisation Schools of Medical Faculties (Milano and Roma). He has been Elected Member of IEEE-EMBS AdCom (Region 8) in the period 1993-1996. He is actually Fellow Member of IEEE and of EAMBES and Associate Editor of IEEE Trans BME. He is a member of the Steering Committee of the IEEE-EMBS Summer School on Biomedical Signal Processing: he was the local organiser of four Summer Schools held in Siena. He has been Visiting Professor at Harvard-MIT Division Health Science and Technology, Boston, USA for an overall period of 1 year. He is the Author of more than 400 international scientific contributions (more than 180 on indexed scientific journals).

1 INTRODUCTION

Biomedical signals and imaging carry important information about the behavior of the living systems under studying. A proper processing of these signals and images allow in many instances to obtain useful physiological and clinical information. Actually, many advanced algorithms of digital signal and image processing are at disposal and therefore

important selective information is now obtainable even in presence of strong sources of noise or low signal/noise ratio. In most of the cases it is not sure whether such sources might derive even by complex and unknown interactions with other biological systems whose implications could be important from the physiological or clinical standpoints. Traditional stationary signal analysis together with innovative methods of investigation of dynamical properties of biological systems and signals in second-order or in higher-order approaches (i.e., in time-frequency, time-variant and time-scale analysis, as well as in non linear dynamics analysis) provide a wide variety of even complex processing tools for information enhancement procedures in the challenging studying of a better explanation of many physiological and clinical phenomena.

2 INTEGRATION BETWEEN SIGNAL PROCESSING AND PHYSIOLOGICAL MODELING

Another important innovative aspect to improve the information content from biomedical data is constituted by the integration between signal processing and modeling of the relevant biological systems, thus directly attributing patho-physiological meaning to the model parameters obtained from the processing; and, viceversa, the modeling fitting could certainly be improved by taking into account the results from signal/image processing procedures.

3 MONOVARIATE AND MULTIVARIATE SIGNAL PROCESSING

Other kinds of integration may be fulfilled, taking into account more signals from the same system in a multivariate way (i.e. from a single-lead vs multichannel EEG or ECG analysis) and combining also the action of different systems such as autonomic nervous system, cardiovascular and respiratory systems, etc. Sleep is a formidable example of multiorgan involvement in both physiological (sleep staging and correlation with cardiorespiratory system) and pathological conditions (sleep apnea, sleep deprivation, restless leg syndrome and so on).

4 MULTISCALE APPROACH

Further, modern rehabilition techniques (motor and /or cognitive) make use actually of objective indices obtained from the patient's biosignals and images to better "personalize" rehabilitation protocols (from EEG, EP's, ERP's, MRI, fMRI, NIRS, etc). In neurosciences such an integration process could comprehend parameters and observations detected also at different scales, from genome and proteome up to the single organ and to the entire body compartment. Examples will be described where an animal model (murine model) is developed by alterating a gene putative to a determined pathology (i.e.epilepsy) and changes in EEG signals are studied (spike/wave occurrences and modifications in signal power bands). In clinical applications, it is worth mentioning the important data fusion which could be fulfilled by the integration of simultaneous EEG recordings and fMRI in some epileptic patients during inter-critical or critical events.

Finally, another important integration can be obtained along different observation scales. Traditionally, biological signal analysis is carried out at the level of organ or system to be investigated (i.e., ECG or EEG signal, arterial blood pressure, respiration and so on). It is very clear the advantage of correlating this information with that one obtained about the same system, but at different scale level, i.e. at cellular level or even at subcellular level (for example, analyzing possible genetic correlates or typical patterns of proteins or even DNA/RNA sequences). Biomedical engineering as a dedicated discipline may strongly contribute to this multiscale information processing

Along this approach line, even the long-QT syndrome, can be efficiently studied at different scale level: a mutation in a portion of gene SCN5A which presents a phenotype compatible to long-QT3 type, is known to produce an altered function of Na+ channels. Through a proper model which describes the functioning of ventricular cells is possible to evidence that this alteration may induce a prolongation of QT duration, as detected on ECG tracing. This event is further correlated with an increased risk of ventricular tachyarrhythmias. Hence, the path is completed: from the genetic expression up to the disease manifestation (Clancy and Rudy, 1999), (Priori ey al., 2003). Many different signal processing and modeling are involved in this paradigmatic example: an integration along the various scales of observation may undoubtedly contribute to a better

understanding of the complex pathophysiological correlates.

A great effort is on course nowadays for creating very large databases and networking of models and technologies for integrating such information (Physiome project (Hunter et al., 2002), (Rudy, 2000) to be connected with Genome and Proteome projects and Virtual Physiological Human project – VPH – which is inserted into the activities of the 7th Framework Programme of EU).

Other examples are constituted by the studying of the profile of expressed proteins in 2D-gel supports, or after mass-spectrometry analysis, relative to a variety of pathologies (i.e. epilepsy, peripheral neuropathies or Amyotrophic Lateral Sclerosis (ALS), or in oncological studies) thus singling out the set of proteins which present a correlate with the pathology in respect to the control group.

This overall approach is reputed promising for obtaining an olistic view of the patient rather than an atomistic one which considers the whole as a simple sum of the single component parts.

REFERENCES

G. Baselli, S. Cerutti, S. Civardi, A. Malliani and M. Pagani, Cardiovascular variability signals: towards the identification of a closed-loop model of the neural control mechanisms, *IEEE Trans Biomed Eng*., vol.35(12):1033-46(1988).

E. N. Bruce, *Biomedical Signal Processing and Signal Modelling*, J. Wiley (2001).

C. E. Clancy and Y. Rudy, Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia, *Nature*, vol. 400(6744):566-569 (1999).

A. Cohen, *Biomedical Signal Analysis*, Vol. I, II, CRC Press (1986).

P. Hunter, P. Robbins and D. Noble, The IUPS human physiome project, *Pflugers Arch-Eur. J. Physiol*., vol. 445:1-9 (2002).

R. I. Kitney, A nonlinear model for studing oscillations in the blood pressure control system*, J. Biomed. Eng*., vol I,n. 2: 88-89 (1979).

H. P. Koepchen, History of studies and concepts of blood pressure wave, in: *Mechanisms of blood pressure waves (*Miyakawa K., Koepchen, H.P., Polosa C. eds): 3-23, Springer-Verlag (1984).

M. C. Mackey and J.C. Milton, Dynamical diseases, *Ann N Y Acad.Sc.,* vol.504:16-32 (1987).

C.-K. Peng, S. Havlin, J.M. Hausdorff, J.E. Mietus, H.E. Stanley and A.L. Goldberger, Fractal mechanisms and heart rate dynamics long-range correlation and their breakdown with diseases, *Journal of Electrocardiology*, vol.28: 59-65 (1995).

S. G. Priori, P.J. Schwartz, C. Napolitano, R. Bloise, E. Ronchetti, M. Grillo, A. Vicentini, C. Spazzolini, J. Nastoli, G. Bottelli, R. Folli and D. Cappelletti, Risk stratification in the long-QT sindrome, *N Engl J Med*., May 8, vol.348(19):1866-74 (2003).

Y. Rudy, From Genome to Physiome: integrative models of cardiac excitation, *Ann. Biom. Engineer*, vol 28: 945-950 (2000).

Cerutti S, Esposti F, Ferrario M, Sassi R, Signorini MG, Long-term invariant parameters obtained from 24-h Holter recordings: a comparison between different analysis techniques, Chaos, vol.17(1):015108, (2007).

Foffani G, Ardolino G, Rampini P, Tamma F, Caputo E, Egidi M, Cerutti S, Barbieri S, Priori A., Physiological recordings from electrodes implanted in the basal ganglia for deep brain stimulation in Parkinson's disease: the relevance of fast subthalamic rhythms, Acta Neurochir Suppl, vol. 93:97-9, (2005).

OUTTHINKING AND ENHANCING BIOLOGICAL BRAINS

Kevin Warwick
University of Reading, UK

Keywords: Brain-Computer Interface, Biological systems, Implant technology, Feedback control.

Abstract: In this paper an attempt has been made to take a look at how the use of implant and electrode technology can now be employed to create biological brains for robots, to enable human enhancement and to diminish the effects of certain neural illnesses. In all cases the end result is to increase the range of abilities of the recipients. An indication is given of a number of areas in which such technology has already had a profound effect, a key element being the need for a clear interface linking the human brain directly with a computer. An overview of some of the latest developments in the field of Brain to Computer Interfacing is also given in order to assess advantages and disadvantages. The emphasis is clearly placed on practical studies that have been and are being undertaken and reported on, as opposed to those speculated, simulated or proposed as future projects. Related areas are discussed briefly only in the context of their contribution to the studies being undertaken. The area of focus is notably the use of invasive implant technology, where a connection is made directly with the cerebral cortex and/or nervous system.

Tests and experimentation which do not involve human subjects are invariably carried out *a priori* to indicate the eventual possibilities before human subjects are themselves involved. Some of the more pertinent animal studies from this area are discussed including our own involving neural growth. The paper goes on to describe human experimentation, in which neural implants have linked the human nervous system bi-directionally with technology and the internet. A view is taken as to the prospects for the future for this implantable computing in terms of both therapy and enhancement.

BRIEF BIOGRAPHY

Kevin Warwick is Professor of Cybernetics at the University of Reading, England, where he carries out research in artificial intelligence, control, robotics and cyborgs. He is also Director of the University KTP Centre, which links the University with Small to Medium Enterprises and raises £2.5 million each year in research income. As well as publishing 500 research papers, Kevin is perhaps best known for his experiments into implant technology. He has been awarded higher doctorates (DScs) both by Imperial College and the Czech Academy of Sciences, Prague. He was presented with The Future of Health Technology Award in MIT, was made an Honorary Member of the Academy of Sciences, St. Petersburg and in 2004 received The IEE Achievement Medal.

1 INTRODUCTION

Research is being carried out in which biological signals of some form are measured, are acted upon by some appropriate signal processing technique and are then employed either to control a device or as an input to some feedback mechanism (Penny et al., 2000), (Roitberg, 2005). In many cases neural signals are employed, for example Electroencephalogram (EEG) signals can be measured externally to the body, using externally adhered electrodes on the scalp (Wolpaw et al., 1990) and can then employed as a control input. Most likely this is because the procedure is relatively simple from a research point of view and is not particularly taxing on the researchers involved. However, reliable interpretation of EEG data is extremely complex – partly due to both the compound nature of the multi-neuronal signals being measured and the difficulties in recording such highly attenuated

In the last few years interest has also grown in the use of real-time functional Magnetic Resonance

Imaging (fMRI) for applications such as computer cursor control. This typically involves an individual activating their brain in different areas by reproducible thoughts (Warwick, 2007) or by recreating events (Pan et al., 2007). Alternatively fMRI and EEG technologies can be combined so that individuals can learn how to regulate Slow Cortical Potentials (SCPs) in order to activate external devices (Hinterberger et al., 2005). Once again the technology is external to the body. It is though relatively expensive and cumbersome.

It is worth noting that external monitoring of neural signals, by means of either EEG analysis or indeed fMRI, leaves much to be desired. Almost surely the measuring technique considerably restricts the user's mobility and, as is especially the case with fMRI, the situation far from presents a natural or comfortable setting. Such systems also tend to be relatively slow, partly because of the nature of recordings via the indirect connection, but also because it takes time for the individual themselves to actually initiate changes in the signal. As a result of this, distractions, both conscious and sub-conscious, can result in false indicators thus preventing the use of such techniques for safety critical, highly dynamic and, to be honest, most realistic practical applications. Despite this, the method can enable some individuals who otherwise have extremely limited communication abilities to operate some local technology in their environment, and, in any case, it can serve as a test bed for a more direct and useful connection.

The definition of what constitutes a Brain-Computer Interface (BCI) is extremely broad. A standard keyboard could be so regarded. It is clear however that various wearable computer techniques and virtual reality systems, e.g. glasses containing a miniature computer screen for a remote visual experience (Mann, 1997), are felt by some researchers to fit this category. Although it is acknowledged that certain body conditions, such as stress or alertness, can be monitored in this way, the focus of this paper is on bidirectional BCIs and is more concerned with a direct connection between a biological brain and technology, and ultimately a human and technology.

2 *IN VIVO* STUDIES

Non-human animal studies can be considered to be a pointer for what is potentially achievable with humans in the future. As an example, in one particular animal study the extracted brain of a lamprey, retained in a solution, was used to control the movement of a small wheeled robot to which it was attached (Reger et al., 2000). The lamprey innately exhibits a response to light reflections on the surface of water by trying to align its body with respect to the light source. When connected into the robot body, this response was utilised by surrounding the robot with a ring of lights. As different lights were switched on and off, so the robot moved around its corral, trying to position itself appropriately.

Meanwhile in studies involving rats, a group of rats were taught to pull a lever in order to receive a suitable reward. Electrodes were then chronically implanted into the rats' brains such that the reward was proffered when each rat thought (one supposes) about pulling the lever, but before any actual physical movement occurred. Over a period of days, four of the six rats involved in the experiment learned that they did not in fact need to initiate any action in order to obtain a reward; merely thinking about it was sufficient (Chapin, 2004).

In another series of experiments, implants consisting of microelectrode arrays have been positioned into the frontal and parietal lobes of the brains of two female rhesus macaque monkeys. Each monkey learned firstly how to control a remote robot arm through arm movements coupled with visual feedback, and it is reported that ultimately one of the monkeys was able to control the arm using only brain derived neural signals with no associated physical movement. Notably, control signals for the reaching and grasping movements of the robotic arm were derived from the same set of implanted electrodes (Carmena et al., 2003), (Nicolelis et al., 2000).

Such promising results from animal studies have given the drive towards human applications a new impetus.

3 ROBOT WITH A BIOLOGICAL BRAIN

Human concepts of a robot may involve a little wheeled device, perhaps a metallic head that looks roughly human-like or possibly a biped walking robot. Whatever the physical appearance our idea tends to be that the robot might be operated remotely by a human, or is being controlled by a simple programme, or even may be able to learn with a

microprocessor/computer as its brain. We regard a robot as a machine.
In a present project neurons are being cultured in a laboratory in Reading University to grow on and interact with a flat multi-electrode array. The neural culture, a biological brain, can be electronically stimulated via the electrodes and its trained response can be witnessed.

The project now involves networking the biological brain to be part of a robot device. In the first instance this will be a small wheeled robot. The input (sensory) signals in this case will be only the signals obtained from the wheeled robot's ultrasonic sensors. The output from the biological brain will be used to drive the robot around. The goal of the project initially will be to train the brain to drive the robot forwards without bumping into any object. Secondly, a separate biological brain will be grown to be the thinking process within a robot head (called Morgui) which houses 5 separate sensory inputs.

What this means is that the brain of these robots will shortly be a biological brain, not a computer. All the brain will know is what it perceives from the robot body and all it will do will be to drive the robot body around or control the robot head respectively. The biological brain will, to all intents and purposes, be the brain of the robot. It will have no life, no existence outside its robotic embodiment.

Clearly this research alters our concept of what a robot is, particularly in terms of ethical and responsibility issues. If a role of animal research is to open up possibilities for future human trials, then in this case the research could well be opening a window on the ultimate possibility of human neurons being employed in a robot body. All the 'human' brain would know would be its life as a robot.

4 HUMAN APPLICATION

At the present time the general class of Brain-Computer Interfaces (BCIs) for humans, of one form or another, have been specifically developed for a range of applications including military weapon and drive systems, personnel monitoring and for games consoles. However, by far the largest driving force for BCI research to date has been the requirement for new therapeutic devices such as neural prostheses.

The most ubiquitous sensory neural prosthesis in humans is by far the cochlea implant (Fin and LoPresti, 2003). Here the destruction of inner ear hair cells and the related degeneration of auditory nerve fibres results in sensorineural hearing loss. As such, the prosthesis is designed to elicit patterns of neural activity via an array of electrodes implanted into the patient's cochlea, the result being to mimic the workings of a normal ear over a range of frequencies. It is claimed that some current devices restore up to approximately 80% of normal hearing, although for most recipients it is sufficient that they can communicate to a respectable degree without the need for any form of lip reading. The typically modest success of cochlea implantation is related to the ratio of stimulation channels to active sensor channels in a fully functioning ear. Recent devices consist of up to 32 channels, whilst the human ear utilises upwards of 30,000 fibres on the auditory nerve. There are now reportedly well over 10,000 of these prostheses in regular operation.

Studies investigating the integration of technology with the human central nervous system have varied from merely diagnostic to the amelioration of symptoms (Warwick and Gasson, 2004). In the last few years some of the most widely reported research involving human subjects is that based on the development of an artificial retina (Rizzo, 2001). Here, small electrode arrays have been successfully implanted into a functioning optic nerve. With direct stimulation of the nerve it has been possible for the otherwise blind recipient to perceive simple shapes and letters. The difficulties with restoring sight are though several orders of magnitude greater than those of the cochlea implant simply because the retina contains millions of photodetectors that need to be artificially replicated. An alternative is to bypass the optic nerve altogether and use cortical surface or intracortical stimulation to generate phosphenes (Dobelle, 2000).

Most invasive BCIs monitor multi-neuronal intracortical action potentials, requiring an interface which includes sufficient processing in order to relate recorded neural signals with movement intent. Problems incurred are the need to position electrodes as close as possible to the source of signals, the need for long term reliability and stability of interface in both a mechanical and a chemical sense, and adaptivity in signal processing to deal with technological and neuronal time dependence. However, in recent years a number of different collective assemblies of microelectrodes have been successfully employed both for recording and stimulating neural activity. Although themselves of small scale, nevertheless high density

connectors/transmitters are required to shift the signals to/from significant signal processing and conditioning devices and also for onward/receptive signal transmission.

Some research has focussed on patients who have suffered a stroke resulting in paralysis. The most relevant to this paper is the use of a '3rd generation' brain implant which enables a physically incapable brainstem stroke victim to control the movement of a cursor on a computer screen (Kennedy, 2000), (Kennedy, 2004). Functional Magnetic Resonance Imaging (fMRI) of the subject's brain was initially carried out to localise where activity was most pronounced whilst the subject was thinking about various movements. A hollow glass electrode cone containing two gold wires and a neurotrophic compound (giving it the title 'Neurotrophic Electrode') was then implanted into the motor cortex, in the area of maximum activity. The neurotrophic compound encouraged nerve tissue to grow into the glass cone such that when the patient thought about moving his hand, the subsequent activity was detected by the electrode, then amplified and transmitted by a radio link to a computer where the signals were translated into control signals to bring about movement of the cursor. With two electrodes in place, the subject successfully learnt to move the cursor around by thinking about different movements. Eventually the patient reached a level of control where no abstraction was needed – to move the cursor he simply thought about moving the cursor. Notably, during the period that the implant was in place, no rejection of the implant was observed; indeed the neurons growing into the electrode allowed for stable long-term recordings.

Electronic neural stimulation has proved to be extremely successful in other areas, including applications such as the treatment of Parkinson's disease symptoms. With Parkinson's Disease diminished levels of the neurotransmitter dopamine cause over-activation in the ventral posterior nucleus and the subthalamic nucleus, resulting in slowness, stiffness, gait difficulties and hand tremors. By implanting electrodes into the subthalamic nucleus to provide a constant stimulation pulse, the over activity can be inhibited allowing the patient, to all external intents and purposes, to function normally (Pinter et al., 1999).

5 BRAIN WITHIN A BRAIN

Ongoing research, funded by the UK Medical Research Council, is investigating how the onset of tremors can be accurately predicted such that merely a stimulation current burst is required rather than a constant pulsing (Gasson et al., 2005: pp.16/1-16/4). This has implications for battery inter-recharge periods as well as limiting the extent of in-body intrusive signalling. The deep brain stimulator can be used to collect local field potential (LFP) signals generated by the neurons around the deep brain electrodes (Gasson et al., 2005: pp.16/1-16/4). Determining the onset of events can be investigated by using fourier transforms to transfer the time based signal to a frequency based spectrogram to determine the change in frequency at the critical time period. However, in addition to that, the frequency changes in the period of time immediately prior to the tremor occurrence can give important information.

Fig.1 shows the results of an initial attempt to train an artificial neural network to indicate not only that a Parkinsonian tremor is present but also that one is very likely to occur in the near future. The aim of this research is that, once a reliable predictor has been obtained, the stimulating pulsing will only be enacted when a tremor is predicted, in order to stop the actual physical tremor occurring before it even starts in the first place.

The bottom trace in Fig.1 shows emg (muscular) signals, measured externally, associated with movement due to the tremors. It can be seen that the tremors in this incident actually start at around the 45 to 50 second point. The trace just above this indicates the corresponding electrical data measured as deep brain Local Field Potentials in the Sub-Thalamic Nucleus of the patient involved. It can be witnessed how, in this case, the electrical data takes on a different form (in terms of variance at least) at around the 45 to 50 second point. The four top plots meanwhile indicate the outputs from 4 differently structured artificial neural networks, based on multi-layer perceptrons with different numbers of neurons in the hidden (middle) layer.

It can be seen how, for each network, the output of the network goes high (logic 1) at the 45 to 50 second point, to indicate the presence of a Parkinsonian tremor. This is all well and good, what is important however is that the output of the networks also briefly goes high around the 30

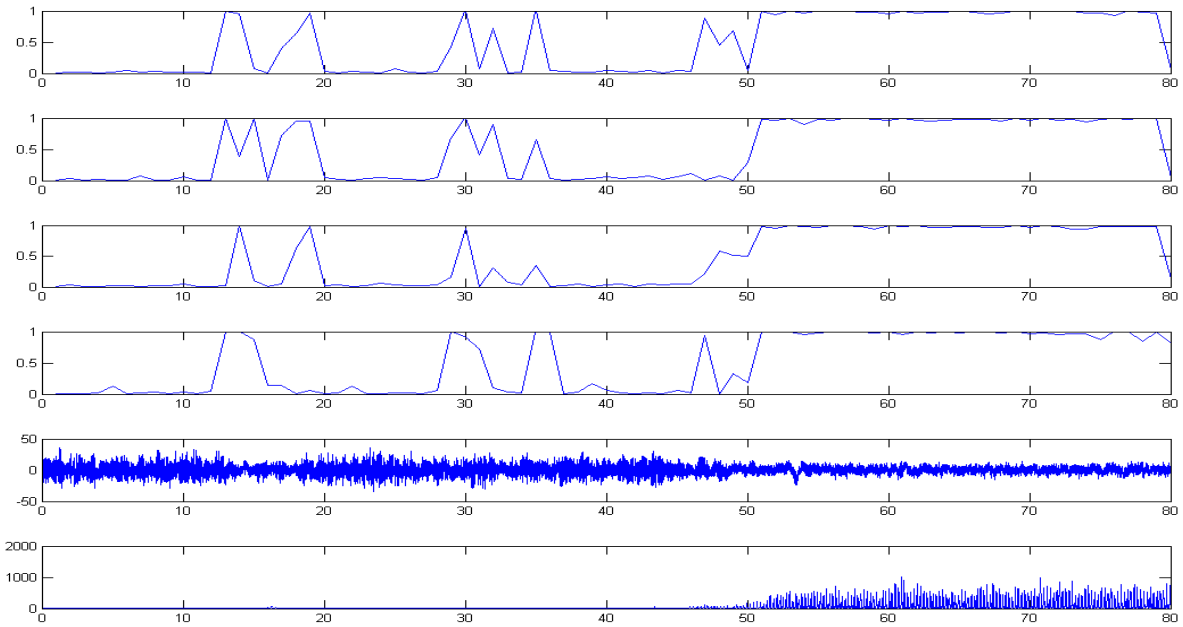


Figure 1: Time plot of the onset of a Parkisonian tremor incident with corresponding artificial neural network indicators.

second point and this can be seen as an indication of the fact that a tremor will shortly occur. Ongoing research is involved with selection of the type and number of inputs to the network, presently these being based on the energy spectrum in different frequency ranges. The networks are also being tested on considerable amounts of resting data, that is long periods of brain activity where no tremors at all actually occur in patients. Clearly the aim is that a network will not give false predictions of tremors.

In fact false positive predictions are not so much of a critical problem. The end result with a false positive is that a stimulation may occur when it is not strictly necessary. In any event no actual tremor would occur, which is indeed a good outcome, however unnecessary energy would have been used – in fact if numerous false predictions occurred the intelligent stimulator would tend toward the present 'blind' stimulator. Effectively the occasional false positive prediction is perhaps not a problem, unless it became a regular occurrence. The good news is that results show that the network can be readily tuned to avoid false positives anyway.

6 GENERAL IMPLANT STUDIES

Some of the most impressive human research to date has been carried out using the microelectrode array, shown in Figure 2. The individual electrodes are only 1.5mm long and taper to a tip diameter of less than 90 microns. Although a number of trials not using humans as a test subject have occurred (Branner and Normann, 2000), human tests are at present limited to two studies. In the second of these the array has been employed in a recording only role (Donoghue et al., 2002), (Donoghue et al., 2004), (Friehs et al., 2004), most notably recently as part of the 'Braingate' system. Essentially activity from a few neurons monitored by the array electrodes is decoded into a signal to direct cursor movement. This has enabled an individual to position a cursor on a computer screen, using neural signals for control combined with visual feedback. The first use of the microelectrode array (Figure 2) will be discussed in the following section as this has considerably broader implications which extend the capabilities of the human recipient.

A key selection point at the present time are what type of implant to employ, as several different possibilities exist, ranging from single electrode devices to multielectrode needles which contain electrode points at different depths to multielectrode arrays which either contain a number of electrodes which penetrate to the same depth (as in Figure 2) or are positioned in a banked/sloped arrangement. A further key area of consideration is the exact positioning of a BCI. In particular certain areas of the brain are, apparently, only really useful for monitoring purposes whilst others are more useful for stimulation.

Actually deriving a reliable command signal from a collection of captured neural signals is not necessarily a simple task, partly due to the complexity of signals recorded and partly due to time constraints in dealing with the data. In some cases however it can be relatively easy to look for and obtain a system response to certain anticipated neural signals – especially when an individual has trained extensively with the system. In fact neural signal shape, magnitude and waveform with respect to time are considerably different to the other signals that it is possible to measure in this situation.

If a greater understanding is required of neural signals recorded, before significant progress can be made, then this will almost surely present a major problem. This is especially true if a number of simultaneous channels are being employed, each requiring a rate of digitization of (most likely) greater than 20KHz in the presence of unwanted noise. For real time use this data will also need to be processed within a few milliseconds (100 milliseconds at most). Further, although many studies have looked into the extraction of command signals (indicating intent) from measured values, it is clear that the range of neural activity is considerable. Even in the motor area not only are motor signals present but so too are sensory, cognitive, perceptual along with other signals, the exact purpose of which is not clear – merely classifying them as noise is not really sufficient and indeed can be problematic when they are repeated and apparently linked in some way to activity.

It is worth stressing here that the human brain and spinal cord are linking structures, the functioning of which can be changed through electronic stimulation such as that provided via an electrode arrangement. This type of technology therefore offers a variety of therapeutic possibilities. In particular the use of implanted systems when applied to spinal cord injured patients, in whom nerve function is disordered, was described in (Warwick, 2004) as having the following potential benefits (among others):

1. Re-education of the brain and spinal cord through repeated stimulation patterns
2. Prevention of spinal deformity
3. Treatment of intractable neurogenic and other pain
4. Assisting bladder emptying
5. Improving bowel function
6. Treatment of spasticity
7. Improvement of respiratory function – assisting coughing and breathing
8. Reduction of cardiovascular maleffects
9. Prevention of pressure sores – possibly providing sensory feedback from denervated areas
10. Improvement and restoration of sexual function
11. Improved mobility
12. Improved capability in daily living, especially through improved hand, upper limb and truncal control

Sensate prosthetics is another growing application area of neural interface technology, whereby a measure of sensation is restored using signals from small tactile transducers distributed within an artificial limb (Fin and LoPresti, 2003). The transducer output can be employed to stimulate the sensory axons remaining in the residual limb which are naturally associated with a sensation. This more closely replicates stimuli in the original sensory modality, rather than forming a type of feedback using neural pathways not normally associated with the information being fed back. As a result it is supposed that the user can employ lower level reflexes that exist within the central nervous system, making control of the prosthesis more subconscious.

One final noteworthy therapeutic procedure is Functional Electrical Stimulation (FES), although it is debatable if it can be truly referred to as a BCI, however it aims to bring about muscular excitation, thereby enabling the controlled movement of limbs. FES has been shown to be successful for artificial hand grasping and release and for standing and walking in quadriplegic and paraplegic individuals as well as restoring some basic body functions such as bladder and bowel control (Grill and Kirsch, 2000). It must be noted though that controlling and coordinating concerted muscle movements for complex and generic tasks such as picking up an arbitrary object is proving to be a difficult, if not insurmountable, challenge.

In the cases described in which human subjects are involved, the aim on each occasion is to either restore functions since the individual has a physical problem of some kind or it is to give a new ability to an individual who has very limited motor abilities. In this latter case whilst the procedure can be regarded as having a therapeutic purpose, it is quite possible to provide an individual with an ability that they have in fact never experienced before. On the one hand it may be that whilst the individual in question has never previously experienced such an ability, some or most other humans have – in this

case it could be considered that the therapy is bringing the individual more in line with the "norm" of human abilities.

It is though also potentially possible to give extra capabilities to a human, to enable them to achieve a broader range of skills – to go beyond the "norm". Apart from the, potentially insurmountable, problem of universally deciding on what constitutes the "norm", extending the concept of therapy to include endowing an individual with abilities that allow them to do things that a perfectly able human cannot do raises enormous ethical issues. Indeed it could be considered that a cochlea implant with a wider frequency response range does just that for an individual or rather an individual who can control the curser on a computer screen directly from neural signals falls into this category. But the possibilities of enhancement are enormous. In the next section we consider how far things could be taken, by referring to relevant experimental results.

7 HUMAN ENHANCEMENT

The interface through which a user interacts with technology provides a distinct layer of separation between what the user wants the machine to do, and what it actually does. This separation imposes a considerable cognitive load upon the user that is directly proportional to the level of difficulty experienced. The main issue it appears is interfacing the human motor and sensory channels with the technology. One solution is to avoid this sensorimotor bottleneck altogether by interfacing directly with the human nervous system. It is certainly worthwhile considering what may potentially be gained from such an invasive undertaking.

Advantages of machine intelligence are for example rapid and highly accurate mathematical abilities in terms of 'number crunching', a high speed, almost infinite, internet knowledge base, and accurate long term memory. Additionally, it is widely acknowledged that humans have only five senses that we know of, whereas machines offer a view of the world which includes infra-red, ultraviolet and ultrasonic. Humans are also limited in that they can only visualise and understand the world around them in terms of a limited dimensional perception, whereas computers are quite capable of dealing with hundreds of dimensions. Also, the human means of communication, essentially transferring an electro-chemical signal from one brain to another via an intermediate, often mechanical medium, is extremely poor, particularly in terms of speed, power and precision. It is clear that connecting a human brain, by means of an implant, with a computer network could in the long term open up the distinct advantages of machine intelligence, communication and sensing abilities to the implanted individual.

As a step towards this more broader concept of human-machine symbiosis, in the first study of its kind, the microelectrode array (as shown in Figure 2) was implanted into the median nerve fibres of a healthy human individual (myself) in order to test *bidirectional* functionality in a series of experiments. A stimulation current direct onto the nervous system allowed information to be sent to the user, while control signals were decoded from neural activity in the region of the electrodes (Gasson et al., 2005:pp 365-375), (Warwick et al., 2003).

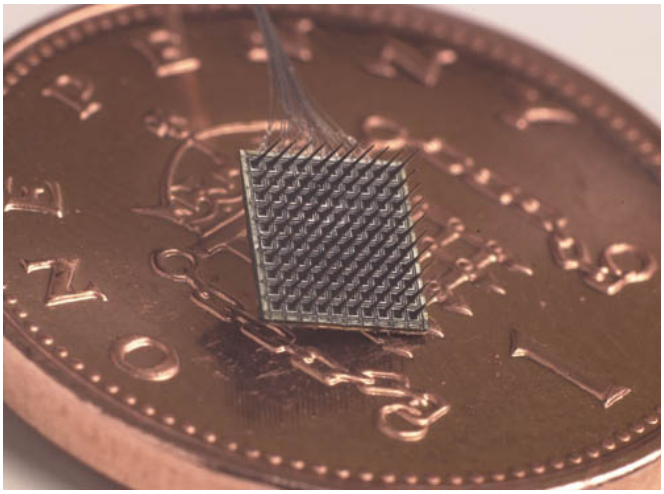


Figure 2: A 100 electrode, 4X4mm Microelectrode Array, shown on a UK 1 pence piece for scale.

In this way a number of experimental trials were successfully concluded (Warwick et al., 2004), (Warwick et al., 2005): In particular:

1. Extra sensory (ultrasonic) input was successfully implemented and made use of.
2. Extended control of a robotic hand across the internet was achieved, with feedback from the robotic fingertips being sent back as neural stimulation to give a sense of force being applied to an object (this was achieved between New York (USA) and Reading(UK))
3. A primitive form of telegraphic communication directly between the nervous systems of two humans was performed.
4. A wheelchair was successfully driven around by means of neural signals.
5. The colour of jewellery was changed as a result of neural signals – as indeed was the behaviour of a collection of small robots.

In each of the above cases it could be regarded that the trial proved useful for purely therapeutic reasons, e.g. the ultrasonic sense could be useful for an individual who is blind or the telegraphic communication could be very useful for those with certain forms of Motor Neurone Disease. However each trial can also be seen as a potential form of augmentation or enhancement for an individual. The question then arises as to how far should things be taken? Clearly enhancement by means of BCIs opens up all sorts of new technological and intellectual opportunities, however it also throws up a raft of different ethical considerations that need to be addressed directly.

8 ON STIMULATION

After extensive experimentation it was found that injecting currents below 80µA onto the median nerve fibers had little perceivable effect. Between 80µA and 100µA all the functional electrodes were able to produce a recognizable stimulation, with an applied voltage of 40 to 50 volts, dependant on the series electrode impedance. Increasing the current above 100µA had no apparent additional effect; the stimulation switching mechanisms in the median nerve fascicle exhibited a non-linear thresholding characteristic.

During this experimental phase, it was pseudo randomly decided whether a stimulation pulse was applied or not. The volunteer (myself), wearing a blindfold, was unaware of whether a pulse had been applied or not, other than by means of its effect in terms of neural stimulation. The user's accuracy in distinguishing between an actual pulse and no pulse at a range of amplitudes is shown in Figure 3.

In all subsequent successful trials, the current was applied as a bi-phasic signal with pulse duration of 200 µsec and an inter-phase delay of 100 µsec. A typical stimulation waveform of constant current being applied to one of the MEA's implanted electrodes is shown in Fig 4.

It was, in this way, possible to create alternative sensations via this new input route to the nervous system. Of the 5 enhancement features mentioned in the previous section, this one will be described, as an example, in further detail. Background information on the other enhancements can be found in a number of references, e.g. (Gasson et al., 2005:pp 365-375), (Warwick et al., 2003), (Warwick et al., 2004), (Warwick and Gasson, 2004).

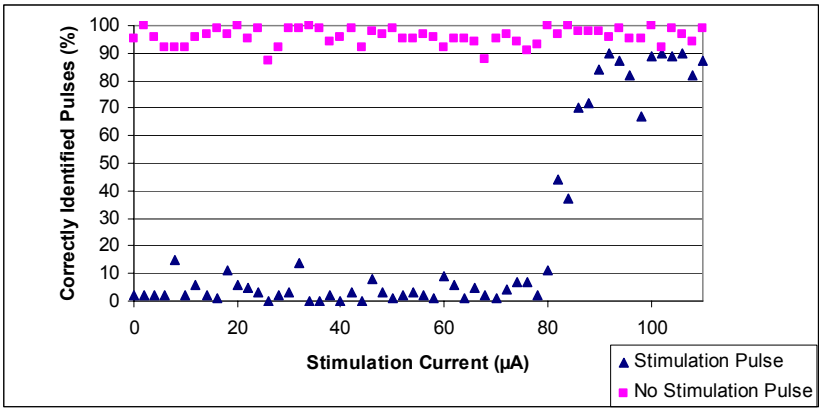


Figure 3: Effect of stimulation amplitude on the number of correctly identified pulses and absence of pulses (over 100 trials).

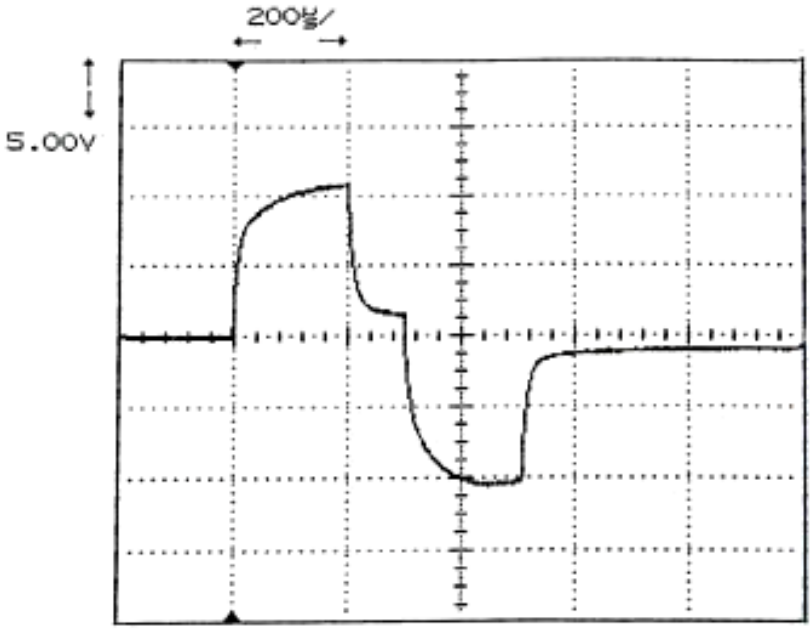


Figure 4: Voltage profile during one bi-phasic stimulation pulse cycle with a constant current of 80µA.

It must be reported that it took 6 weeks for my brain to repetitively recognize the stimulating signals accurately. This time period can be due to a number of contributing factors:

(a) The team had to learn which signals (what amplitude, frequency etc.) would be best in order to bring about a recognizable stimulation.

(b) The recipient's brain had to learn to recognize the new signals it was receiving.

(c) The bond between the recipient's nervous system and the implant was physically changing (becoming stronger).

9 EXTRA SENSORY EXPERIMENT

An experiment was set up to determine if the human brain is able to understand and successfully operate with sensory information to which it had not previously been exposed. Whilst it is quite possible to feed in such sensory information via a normal human sensory route, e.g. electromagnetic radar or infra-red signals are converted to visual, what we were interested in was feeding such signals directly onto the human nervous system, thereby bi-passing the normal human sensory input.

Ultrasonic sensors were fitted to the rim of a baseball cap (see Figure 5) and the output from these sensors, in the form of a proportional count, was employed to bring about a direct stimulation of the nervous system. Hence when no objects were in the vicinity of the sensors, no stimulation occurred, and as an object moved close by so the rate of stimulation pulses being applied increased in a linear fashion up to a pre-selected maximum rate. No increase in stimulation occurred when an object moved closer than 10cm to the sensors.

The ultrasonic sensors were open type piezoelectric ceramic transducers with conical metal resonators and operated at 40 KHz. These were used in a pair, one for transmit and one for receive, to give maximum sensitivity for small and distant objects. The most useful range for the experimentation was found to be 2 – 3m, this being also dependent on the size of object. A simple microcontroller was programmed to perform the echo ranging on the pair of transducers, and provide the range to the first detectable object only. This was translated into a stimulation pulse train, which operated on a single pin of the electrode array. Pins on the array had been tested for their suitability for stimulation by the earlier experimentation in which the recipient identified the presence or absence of stimulation pulse trains at various amplitudes and repetition frequencies.

Figure 5: Experimentation and testing of the ultrasonic baseball cap.

It was found that very little learning was required for the new ultrasonic sense to be used effectively and successfully – merely a matter of 5/6 minutes. This said it must be remembered that it had already taken several weeks for the recipient's brain to successfully, accurately recognize the current signals being injected.

As a result, in a witnessed experiment, the recipient, whilst wearing a blindfold, was able to move around successfully within a cluttered laboratory environment, albeit at a slower than normal walking pace. The sensory input was "felt" as a new form of sensory input (not as touch or movement) in the sense that the brain made a direct link between the signals being witnessed and the fact that these corresponded in a linear fashion to a nearby object.

10 CONCLUSIONS

External input-output interfaces with human and animal brains have been studied for many years. These are sometimes referred to as Brain-Computer Interfaces (BCIs) even though the interface may be external to the (human) body and its sensorimotor mechanism. In this paper an attempt has been made to put such systems in perspective. Emphasis has been placed on such interfaces that can be obtained by means of implanted devices through invasive surgery and actual direct neural connections. In particular a number of trials in this area have clearly shown the possibilities of monitoring, stimulating and enhancing brain functioning.

Although there is no distinct dividing line it is quite possible as far as humans are concerned to investigate BCIs in terms of those employed for direct therapeutic means and those which can have an enhanced role to play. It is clear that the interaction of electronic signals with the human brain can cause the brain to operate in a distinctly different manner. Such is the situation with the stimulator implants that are successfully used to counteract, purely electronically, the tremor effects associated with Parkinson's disease. Such technology can though potentially be employed to modify the normal functioning of the human brain and nervous system in a number of different ways.

The same stimulator, with slightly different positioning, has been shown to elicit feelings of sadness or happiness in the recipient. Given the nature of the intelligent stimulator described here it would appear to be possible to monitor, in real time, a human brain with a computer brain, and for the computer brain to predict when the human is going to feel sad – quite some time before they actually feel sad. In theory a signal could then be injected at

that time to make them feel happy, or at least to stop them actually ever feeling sad in the first place. Maybe this could be regarded as an electronic anti-depressant. There are of course questions about recreational use here – but this would need a deep brain implant which might well prove to be rather too onerous for most people.

Perhaps understandably, invasive BCIs are presently far less well investigated in University experiments than their external BCI counterparts. A number of animal trials have though been carried out and the more pertinent have been indicated here along with the relevant human trials and practice. In particular the focus of attention has been given to the embodiment of grown neural tissue within a technological body. Whilst only 1,000 or so neurons are involved this presents an interesting research area in a number of ways. But once the number of such neurons used increases 1,000 or 1,000,000-fold, it also raises enormous philosophical and ethical issues. For example is the robot 'thinking' and what rights should it have?

The potential for BCI applications for individuals who are paralysed is enormous, where cerebral functioning to generate command signals is functional despite the motor neural pathways being in some way impaired – such as in Lou Gehrig's disease. The major role is then either one of relaying a signal of intention to the appropriate actuator muscles or to reinterpret the neural signals to operate technology thereby acting as an enabler. In these situations no other medical 'cure' is available, something which presents a huge driver for an invasive implant solution for the millions of individuals who are so affected. Clearly though, bidirectional signalling is important, not only to monitor and enact an individual's intent but also to provide feedback on that individual's resultant interaction with the real world. For grasping, walking and even as a defensive safety stimulant, feedback is vital. This paper has therefore focussed on such studies.

Where invasive interfaces are employed in human trails, a purely therapeutic scenario often exists. In a small number of instances, such as use of the microelectrode array as an interface, an individual has been given different abilities, something which opens up the possibilities of human enhancement. These latter cases however raise more topical ethical questions with regard to the need and use of a BCI. What might be seen as a new means of communication for an individual with an extreme form of paralysis or a new sensory input for someone who is blind, opening up a new world for them, can also be seen as an unnecessary extra for another individual, even though it may provide novel commercial opportunities. What is therapy for one person may be regarded as an enhancement or upgrading for another.

Whilst there are still many technical problems to be overcome in the development of BCIs, significant recent experimental results have indicated that a sufficient technological infrastructure now exists for further major advances to be made. Although a more detailed understanding of the underlying neural processes will be needed in the years ahead, it is not felt that this will present a major hold up over the next few years, rather it will provide an avenue of research in which many new results will shortly appear through trials and experimentation, possibly initially through animal studies although it must be recognised that it is only through human studies that a full analysis can be made and all encompassing conclusions can be drawn. Nevertheless the topic opens up various ethical questions that need to be addressed and as such, research in this area should, I believe, only proceed in light of a pervasive ethical consensus.

ACKNOWLEDGEMENTS

The Author would like to acknowledge the considerable assistance and input of the Consultant Neurosurgeons Mr. Peter Teddy, Mr. Amjad Shad, Mr. Ali Jamous and Mr. Tipu Aziz and researchers Iain Goodhew, Mark Gasson, Ben Whalley and Ben Hutt. Ethical approval for the author's research was obtained from the Ethics and Research Committee at the University of Reading, UK and with regard to the neurosurgery aspect, the Oxfordshire National Health Trust Board overseeing the Radcliffe Infirmary, Oxford, UK.

REFERENCES

Branner, A. and Normann, R. (2000) A multielectrode array for intrafascicular recording and stimulation in the sciatic nerve of a cat, *Brain Research Bulletin*, Vol.51, pp.293-306.

Chapin, J.K. (2004) Using multi-neuron population recordings for neural prosthetics. *Nature Neuroscience*, 7, 452-454.

Carmena, J., Lebedev, M., Crist, R., O'Doherty, J., Santucci, D., Dimitrov, D. Patil, P., Henriquez, C. and Nicolelis, M. (2003) Learning to control a brain-

machine interface for reaching and grasping by primates, *Plos Biology*, Vol.1, Issue.2, article number e2.
Dobelle, W. (2000) Artificial vision for the blind by connecting a television camera to the visual cortex, ASAIO J, Vol.46, pp.3-9.
Donoghue, J. (2002) Connecting cortex to machines: recent advances in brain interfaces, *Nature Neuroscience Supplement*, Vol.5, pp.1085-1088.
Donoghue, J., Nurmikko, A., Friehs, G. And Black, M. (2004) Development of a neuromotor prosthesis for humans, Chapter 63 in *Advances in Clinical Neurophysiology, Supplements to Clinical Neurophysiology*, Vol.57, pp.588-602.
Finn, W. and LoPresti, P. (eds.) (2003) Handbook of Neuroprosthetic methods, CRC Press.
Friehs, G., Zerris, V., Ojakangas, C., Fellows, M. and Donoghue, J. (2004) Brain-machine and brain-computer interfaces, *Stroke*, Vol.35, Issue.11, pp.2702-2705.
Gasson, M., Hutt, B., Goodhew, I., Kyberd, P. and Warwick, K. (2005) Invasive neural prosthesis for neural signal detection and nerve stimulation, Proc. International Journal of Adaptive Control and Signal Processing, Vol.19, No.5, pp.365-375.
Gasson, M., Wang, S., Aziz, T., Stein, J. and Warwick, K. (2005) Towards a demand driven deep brain stimulator for the treatment of movement disorders, Proc. 3rd IEE International Seminar on Medical Applications of Signal Processing, pp.16/1-16/4.
Grill, W. and Kirsch, R. (2000) Neuroprosthetic applications of electrical stimulation, *Assistive Technology,* Vol.12, Issue.1, pp.6-16.
Hinterberger, T., Veit, R., Wilhelm, B., Weiscopf, N., Vatine, J. and Birbaumer, N. (2005) Neuronal mechanisms underlying control of a brain-computer interface, *European Journal of Neuroscience*, Vol.21, Issue.11, pp.3169-3181.
Kennedy, P., Bakay, R., Moore, M., Adams, K. and Goldwaith, J. (2000) Direct control of a computer from the human central nervous system, *IEEE Transactions on Rehabilitation Engineering*, Vol.8, pp.198-202.
Kennedy, P., Andreasen, D., Ehirim, P., King, B., Kirby, T., Mao, H. and Moore, M. (2004) Using human extra-cortical local field potentials to control a switch, *Journal of Neural Engineering*, Vol.1, Issue.2, pp.72-77.
Mann, S. (1997) Wearable Computing: A first step towards personal imaging, *Computer*, Vol. 30, Issue.2, pp. 25-32.
Nicolelis, M., Dimitrov, D., Carmena, J., Crist, R., Lehew, G., Kralik, J. and Wise, S. (2003) Chronic, multisite, multielectrode recordings in macaque monkeys, *Proc. National Academy of the USA*, Vol.100, Issue.19, pp.11041-11046.
Penny, W., Roberts, S., Curran, E., and Stokes, M. (2000) EEG-based communication: A pattern recognition approach, *IEEE Transactions on Rehabilitation Engineering*., Vol. 8, Issue.2, pp. 214-215.
Pinter, M., Murg, M., Alesch, F., Freundl, B., Helscher, R. And Binder, H. (1999) Does deep brain stimulation of the nucleus ventralis intermedius affect postural control and locomotion in Parkinson's disease?, *Movement Disorders*, Vol.14, Issue.6, pp.958-963.
Reger, B., Fleming, K., Sanguineti, V., Simon Alford, S., Mussa-Ivaldi, F. (2000) Connecting Brains to Robots: an artificial body for studying computational properties of neural tissues, *Artificial life*, Vol.6, Issue.4, pp.307-324.
Rizzo, J., Wyatt, J., Humayun, M., DeJuan, E., Liu, W., Chow, A., Eckmiller, R., Zrenner, E., Yagi, T. and Abrams, G. (2001) Retinal Prosthesis: An encouraging first decade with major challenges ahead, *Opthalmology,* Vol.108, No.1.
Roitberg, B. (2005) Noninvasive brain-computer interface, *Surgical Neurology*, Vol.63, Issue.3, p.195.
Warwick, K. (2004) I Cyborg, University of Illinois Press.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Andrews, B., Teddy, P., Shad, A. (2003) The application of implant technology for cybernetic systems. *Archives of Neurology*, 60 (10), pp.1369-1373.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Schulzrinne, H. and Wu, X. (2004) Thought Communication and Control: A First Step Using Radiotelegraphy, *IEE Proceedings on Communications,* Vol.151, No. 3, pp 185-189.
Warwick, K., Gasson, M., Hutt, B. and Goodhew, I. (2005) An Attempt to Extend Human Sensory Capabilities by means of Implant Technology. Proc. IEEE Int. Conference on Systems, Man and Cybernetics, Hawaii.
Wolpaw, J., McFarland, D., Neat, G. and Forheris, C. (1990) An EEG based brain-computer interface for cursor control, *Electroencephalogr. Clin. Neurophysiol*., Vol.78, pp.252-259.
Pan, S., Warwick, K., Gasson, M., Burgess, J., Wang, S., Aziz, T. and Stein, J. (2007) Prediction of parkinson's disease tremor onset with artificial neural networks", Proc. IASTED Conference BioMed 2007, Innsbruck, Austria, pp.341-345.
Warwick, K. (2007) The promise and threat of modern cybernetics, Southern Medical Journal, Vol.100, Issue.1, pp.112-115.
Warwick, K. and Gasson, M.N. (2004) Practical Interface Experiments with Implant Technology, in "Computer Vision in Human-Computer Interaction", Sebe, N., Lew, M.S and Huang, T.S. (eds.), Lecture Notes in Computer Science, Vol. 3058, pp.7-16.Xie, S., Yang, Z. and Yang, Y. (2004) Brain-computer interface based on event-related potentials during imitated natural reading, International Journal of Psychology, Vol.39, Issue.5-6, Supplement S., p.138.
Yoo, S., Fairneny, T., Chen, N., Choo, S., Panych, L., Park, H., Lee, S. and Jolesz, F. (2004) Brain-computer interface using fMRI: spatial navigation by thoughts, *Neuroreport*, Vol.15, Issue.10, pp.1591-1595.
Yu, N., Chen, J. and Ju, M. (2001) Closed-Loop Control of Quadriceps/Hamstring activation for FES-Induced Standing-Up Movement of Paraplegics, *Journal of Musculoskeletal Research*, Vol. 5, No.3, pp.173-184.

ANALYSIS AND MODELS OF BRAIN EPILEPTIC ACTIVITIES

Fernando Henrique Lopes da Silva
University of Amsterdam, The Netherlands

Abstract: The essence of epilepsy is the sudden occurrence of a qualitative change in the behaviour of neuronal networks of some specific areas of the brain. In general we may assume that neuronal networks possess multistable dynamics. We may simplify this concept considering the case that a neuronal network may display, at least, two dynamical states: an interictal state characterised by a normal on-going neural activity, as revealed in the Eletcroencephalogram of Magnetoencephalogram (EEG, MEG), that may be apparently random, and another one – the ictal state - that is characterised by the sudden occurrence of synchronous oscillations, most commonly with large amplitude. The latter becomes manifest as a paroxysmal change of behaviour and /of the state of consciousness of a patient, i.e. an epileptic seizure. In the terminology of the mathematics of non-linear systems, we may state that a neuronal network behaves as a bistable system with two attractors, to which the system converges depending on initial conditions and on the system's parameters.

We propose schematically that the transition between the normal on-going to the seizure activity can take place according to three basic models: Model I – a transition may occur due to random fluctuations of some system's parameters. These transitions are thus unpredictable. Models II and III – a transition may result from a gradual change of some unstable parameters, either due to endogenous (model II) or exogenous (model III). In these cases the change of parameter values causes a deformation of the attractor resulting in a transition from the basin of the attractor corresponding to the normal state, to the attractor corresponding to the seizure dynamical state. Some experimental findings obtained in different cases of epilepsy, both in human and in animals, are compatible with each of these 3 models. Some examples of these cases are illustrated.

BRIEF BIOGRAPHY

Fernando Henrique Lopes da Silva received his Medical Degree from the University of Lisbon in 1959, got his Ph.D. from the University of Utrecht in 1970, and in 1980 was appointed Full Professor of General Physiology at the Faculty of Science at the University of Amsterdam (since 2002 part of the Swammerdam Institute for Life Sciences). From 1993 to 2000 he was Director of the newly created Institute of Neurobiology of the University of Amsterdam, and member of the Scientific Directorate of the Graduate School Neurosciences Amsterdam. In 2000, when he reached the retirement age of 65, he became Emeritus Professor of the same University, and has at present a free-lance contract with the Swammerdam Institute for Life Sciences.

Since 1970, he supervised a large number of student trainees from different Universities and Faculties: Medical, Biology, Sciences, (Bio-medical) Engineering. Supervised 65 Ph.D. students (up to December 2006).

His research interests are centred on the biophysical aspects of electrical activity of the brain and the functional organization of neuronal networks, namely of the cerebral cortex and the limbic system, with a special interest in the generation and functional significance of brain rhythmic activities. He published more than 220 papers in peer-reviewed journals and contributed Chapters to 10 multi-authored books (of 6 he is co-editor), among which the Handbook "Electroencephalography: Basic principles, clinical applications and related fields", Niedermeyer, E. and Lopes da Silva, F.H. (Eds), published by Lippincott, Williams and Wilkins, Baltimore; 5 Editions: 1982, 1987, 1993, 1998, 2004. In addition he contributed chapters to to the Encyclopedia of Neuroscience (George Adelman, Barry H. Smith. Eds), Elsevier Science, 2003 (3rd edition), to the Encyclopedia of the Human Brain (Ed. V. S. Ramachandran), Academic Press, 2002, and to The Handbook of Brain Theory and Neural Networks (Ed. Michael A. Arbib), The MIT Press, 2003 (2nd edition).

Selection of Scientific Awards

- 1975 He received the Winkler Medal from the Netherlands Association for Neurology for scientific contributions in the field of neurosciences.
- 1985 Elected member of the Royal Netherlands Academy of Arts and Sciences.
- 1990 "Lord Adrian" Lecturer at the 12th World Congress of Electroencephalography and Clinical Neurophysiology in Rio de Janeiro, Brazil.
- 1992 Honorary President of the VIIth European Congress of Clinical Neurophysiology, Budapest, Hungary.
- 1995 Honorary Life Member of The British Society for Clinical Neurophysiology (Formerly The EEG Society), London, United Kingdom.
- 1997 Doctor Honoris Causa of the University of Lisbon (Portugal).
- 1997 Special "Berger" Lecturer at the 14th International Congress of EEG and Clinical Neurophysiology in Florence, Italy.
- 1999 Recipient of the Herbert H. Jasper Award, selected by the American Clinical Neurophysiology Society for his "lifetime of outstanding contributions to the field of clinical neurophysiology.''
- 2000 Recipient of the 'Storm van Leeuwen/ Magnus Prize' of the Dutch Society of Clinical Neurophysiology.
- 2000 Honorary member of the Portuguese Society of Electroencephalography and Clinical Neurophysiology.
- 2002 Recipient of the Ragnar Granit Prize for his work on the field of Bioelectromagnetism.
- 2002 Doctor Honoris Causa of the University of Porto (Portugal).
- 2004 Recipient of the first Prize "Universidade de Coimbra" for a (sic) "person of Portuguese nationality who has made a particular relevant and innovative contribution in the fields of culture or science."

General Honors

- 2000 High Officer of the Order of Santiago da Espada, for outstanding achievements in the field of Science/Art/Literature, awarded by the President of the Republic of Portugal.
- 2001 Knight of the Order of the 'Nederlandse Leeuw' awarded by the Queen of the Netherlands in appreciation for his achievements in science.

FROM THE BENCH TO THE BEDSIDE

The Role of Semantics in Enabling the Vision of Translational Medicine

Vipul Kashyap
Partners HealthCare System, Clinical Informatics R&D, USA

Abstract: Biomedical research and healthcare clinical transactions are generating huge volumes of data and information. At the same time, the results of biomedical research in the form of new molecular diagnostic tests and therapies are being increasingly used in the context of clinical practice. There is a critical need to speed "translation" of genomic research insights into clinical research and practice. In this talk, we will discuss challenges faced by a healthcare enterprise in realizing the vision of Translational Medicine, such as:

- The need to create structured and semantic representations of genotypic and phenotypic data such as clinical observations and molecular diagnostic tests.
- The need for cost-effective and incremental data integration for combining genotypic and phenotypic information at the point of care.
- The need for actionable decision support for suggesting molecular diagnostic tests and therapies in the context of clinical care.
- The need for knowledge update, propagation and consistency to keep abreast of the rapid pace of knowledge discovery being witnessed in the life sciences, a crucial pre-requisite to reduce the cost of knowledge acquisition and maintenance.

Semantics-based approaches to address the above-mentioned challenges, including the applicability of semantic web standard (RDF, OWL, Rules); and issues related to the value proposition of these technologies will be presented.

BRIEF BIOGRAPHY

Vipul Kashyap, PhD is a Senior Medical Informatician in the Clinical Informatics Research & Development group at Partners HealthCare System and is currently the chief architect of a Knowledge Management Platform that enables browsing, retrieval, aggregation, analysis and management of clinical knowledge across the Partners Healthcare System. Vipul received his PhD from the Department of Computer Science at Rutgers University in New Brunswick in the area of metadata and semantics-based knowledge and information management. He is also interested in characterization of the value proposition of semantic technologies in the enterprise context. Before coming to Partners, Vipul has held positions at MCC, Telcordia (Bellcore) and was a fellow at the National Library of Medicine. Vipul has published 2 books on the topic of Semantics, 40-50 articles in prestigious conferences and journals; and has participated in panels and presented tutorials on the topic of semantic technologies. Vipul sits on the technical advisory board of an early stage company developing semantics-based products, and represents Partners on the W3C advisory committee and the HealthCare Information Technology Standards Panel (HITSP).

THE CANCER INFORMATICS ECOSYSTEM

A Case Study in the Accretion of Federated Systems based on Service Oriented Architectures, Semantic Integration and Computing Grids

David Hall
Research Triangle Institute in North Carolina, USA

Abstract: Information technology is playing an increasingly critical role in health and life sciences research due to the profound expansion in the scope of research projects in the post-genomic age. Robust data management and analysis systems are becoming essential enablers of these studies. Driven by funding agency requirements, funding opportunities, and grass roots organizing, efforts are underway to develop standards and technologies to promote large-scale integration of publicly-funded systems and databases including infrastructure developed for individual studies. Predicted benefits include an enhanced ability to conduct meta-analyses, an increase in the usable lifespan of data, a funding agency-wide reduction in the total cost of IT infrastructure, and an increased opportunity for the development of third party software tools. This presentation will critically examine efforts towards developing publicly-accessible interoperable and distributed production systems in the health and life sciences via ontologies, formal metadata, service oriented architectures, and grid computing models with a focus on several projects under the direction of the author in the area of cancer informatics.

BRIEF BIOGRAPHY

David Hall is a Senior Software Project Leader at RTI International based in North Carolina, USA. He leads teams of up to 30 developers implementing computer systems that support large biomedical and biotechnological research enterprises in cancer research, drug discovery, genetic epidemiology, and plant biotechnology. His area of interest is the practical application of bioinformatics and medical informatics methods, technologies, and standards in the development of production software. Particular topics of interest include data visualization, semantic integration, systems integration, and high performance computing. Recent clients include the US National Institutes of Health, GlaxoSmithKline, Syngenta, and Duke University. Data systems developed by David's group manage clinical and research data for nearly one million patients. Applications include data warehouses, metadata registries, workflow systems, high resolution image databases, analytical applications, and web services. David is currently Principal Investigator of the Informatics Support Center for the National Cancer Institute's Breast and Colon Cancer Family Registries. He holds a Ph.D. in Genetics from the University of Georgia and a B.S. in Computer Science from Wake Forest University.

ICT AND PERSONS WITH DISABILITIES
The Solution or the Problem?

Albert M. Cook
Faculty of Rehabilitation Medicine, University of Alberta, Edmonton Alberta, Canada
al.cook@ualberta.ca

Keywords: Assistive technologies, information and computer technologies, persons with disabilities.

Abstract: In order to lead full and productive lives, persons with disabilities need to have the same access to information and communication systems as the rest of the population. Advances in information and communication technologies (ICT) are occurring quickly, and the capability of technologies to meet the needs of persons with disabilities is growing daily. Future developments in assistive technologies (AT) and the successful application of these technologies to meet the needs of people who have disabilities are dependent on exploitation of these ICT advances. AT also involves the development of specialized interfaces such as the brain computer interface (BCI), adaptive interfaces that accommodate for changes in the user's physical skills, cognitive interfaces that allow understanding of the human technology interface by individuals with intellectual disabilities and systems that accommodate for user needs based on environmental sensing (e.g., GPS interfaces) and downloading of profiles to meet specific user needs. Universal Design (or design for all) calls for the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. In the physical world this often means ramps, curb cuts and other adaptations to the built environment to accommodate individuals who have disabilities. In the ICT world the barriers to access are technological, and the goal for ICT universal design is to have an environment with enough embedded intelligence to be easily adaptable to the varying cognitive, physical and sensory skills of a wide range of individual's in order to meet their productivity, leisure and self care needs. If ICT advances are not adaptable enough to be accessible to persons with disabilities it will further increase the disparity between those individuals and the rest of the population leading to further isolation and economic disadvantage. On the other hand, availability of these technologies in a transparent way will contribute to full inclusion of individuals who have disabilities in the mainstream of society.

BRIEF BIOGRAPHY

Dr. Albert Cook is Professor of Speech Pathology and Audiology and currently Dean of the Faculty of Rehabilitation Medicine and Chair of the Health Sciences Council at the University of Alberta. Dr. Cook has worked with interdisciplinary teams to develop assistive devices and to assess the effectiveness of technology being used by persons with disabilities. Dr. Cook is also associated with the I CAN Centre the Glenrose Rehabilitation Hospital. He was formerly Professor of Biomedical Engineering at California State University, Sacramento where he established the graduate program in biomedical engineering and directed it for 12 years. He also served as Co-Director of the Assistive Device Center in Sacramento, California, helping over 500 persons with disabilities to identify and obtain assistive technologies.

He received his Bachelor of Science in Electrical Engineering at the University of Colorado, a Masters in Bioengineering and his doctorate from the University of Wyoming He is a member of Tau Beta Pi, Phi Kappa Phi and Gold Key honorary societies.

Dr. Cook co-authored with Janice Polgar, OTR, Cook and Hussey's Assistive Technologies: Principles and Practice 3nd edition, published in October 2007 by Elsevier. He has co-edited three other textbooks with John Webster and others and has written numerous chapters in rehabilitation and biomedical engineering texts and monographs.

Dr. Cook's research interests include augmentative and alternative communication, biomedical instrumentation and assistive technology design, development and evaluation. His most recent

research has focussed on the use of robotics with young children who have severe disabilities to develop and assess cognitive and linguistic skills. He has US and foreign patents and numerous publications and conference presentations in these areas. He has been principal investigator on research and training grants in augmentative communication, assistive technologies and biomedical engineering.
Dr. Cook is Past-President and Fellow of RESNA, a major professional society for assistive technology practitioners in North America. He has also served in national united States positions in the Institute of Electrical and Electronic Engineers Engineering in Medicine and Biology Society, the American Society for Engineering Education, the Biomedical Engineering Society, the International Society for Augmentative and Alternative Communication and the Association for the Advancement of Medical Instrumentation. Dr. Cook is a registered professional engineer (electrical) in California.

1 ICT AND PERSONS WITH DISABILITIES TECHNOLOGY AND PROGRESS

Societal Progress requires change much of which is accomplished through advances in technology. In his book, *A Short History of Progress*, Ronald Wright *(*2004) points out that this characteristic has been true for millions of years as societies have advanced through greater utilization of technology.

Wright goes on to describe the problems that technology typically creates such as over consumption, environmental ruin, and separation of classes. These problems are amplified for people who have disabilities, and they lead to a gap in the access to work, self care and community participation for persons with disabilities compared to the general population. Since people with disabilities often depend on technologies for societal participation, the lack of availability of accessible technology or the obsolescence of accessible technologies isolates them further. This is an extension of the concept of the "digital divide" that separates people along socioeconomic lines based on their access to ICT. I refer to it as the "disability gap".

2 ADVANCES IN INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT)

The 21st Century is characterized by a continuous move from a machine-based to a knowledge based economy (Ungson & Trudel, 1999). In this shift, the basis of competence is changing to knowledge skills from machine skills. Information currently amounts to 75% of value added to products This will continually increase, and connectivity will be the key to business success. There is also a move from a regional or national scope of business influence to a global scope, in which virtual networks dictate organizational structures.

Key players in business development are becoming communication suppliers with the move from host-based to network based systems. Telephone, cable TV and internet service providers control commercial growth. Along with these changes networks will become more graphically-based moving increasingly from text-based systems. In order to lead full and productive lives, persons with disabilities need to have the same access to this new information and communication system as the rest of the population.

2.1 What Can we Expect from Technology in the Next 20 Years?

The cost of information technology is continually dropping for comparable or increased computing power and speed. There is also a greater understanding of the biological/physical interface for the control of computers. The human computer interface (HCI) is being developed to be more human-like, more user oriented and more intelligent-providing additional capabilities for searching, processing and evaluating information.

There are a number of changes that are likely to occur over the next few years (Applewhite, 2004). There will be an increase in automated transactions between individuals and organizations enabling people to complete all transactions without face-toface interactions. It is expected that we will achieve equalized access to the web and information between the developed and developing world. Embedded systems will dramatically increase with application such as "intelligence in the doorknob" that recognizes the owner and doesn't require key manipulation. We are likely to see much greater

understanding of the biological to physical interface for the control of computers.

2.2 Changes in Mainstream Tech with AT Implications

There are many examples of emerging mainstream technologies with potential for assisting people with disabilities to access ICT systems. A few of these are described in this section.

Display-based assistive technologies present an array of choices for a user to select from (Cook & Polgar, 2007). This often referred to as scanning since the choices are highlighted sequentially and then chosen using some sort of gross movement. One of the problems associated with this approach is that there must be a physical display for making selections. This often requires the overall system to be larger and more bulky or places a display between a user and a communication partner. A new development is a direct retinal display that creates image that overlays view of real object (Lewis, 2004). The retinal display is low powered because it is shined on retina directly. Scanning light into the eye allows the image to overlay an object such as a communication partner's face-enabling eye contact and small size. The scanning array could be the retinal image, since display scans across the retina power levels can be kept low kept low for safety.

Another development is 3-D displays that create a more intuitive view of objects, events and activities (Lewis, 2004). This type of display may be helpful to individuals who have cognitive disabilities. It might also create new challenges for individuals with visual limitations.

Embedded automatic speech recognition is being developed for PDAs because of the need for keyboards with more and more functions and the limitations of very small keyboards (Kumagai, 2004). This feature could be very useful to reduce individuals who have limited hand function or for those who cannot see the keyboard to make entries.

3 MEETING THE ICT NEEDS OF PERSONS WITH DISABILITIES

Over the centuries, our ability to make tools is what distinguishes us as human, but our tools ultimately control us by making us dependent on them (Wright, 2004). This dependence is less optional for people who have disabilities

3.1 Impact of Technology Advances on People who have Disabilities

Technology advances increase the gap between people who have disabilities and those who don't (Wright, 2004). All societies become hierarchical with an upward concentration of wealth (including aggregations of technology tools) that ensures that "there can never be enough to go around", and this disparity contributes to the "digital divide" and the "disability gap". As advances occur more quickly, the gap widens faster and the people who are poor and/or disabled loose out even more completely and faster. This is a characteristic of cultural and societal "progress" over centuries-technology drives change, and creates both positive and negative outcomes in the process

The prognosis is not good for people with disabilities unless there is considerable effort to keep them connected to ICT and thereby to commerce, employment and personal achievement. There two fundamental approaches to this problem (1) make mainstream technologies accessible to people who have disabilities, or (2) design special purpose technologies specifically for people with disabilities. The former approach is referred to as *universal design* or *design for all*. The second approach utilizes *assistive technologies*.

3.2 Implications for Assistive Technologies

Access to ICT for people with disabilities is a significant global problem, and it has major implications for assistive technologies. There is a constant challenge to keep ICT systems accessible to persons who have disabilities as mainstream advances occur and adaptations become potentially incompatible with the new systems. Communication technologies change rapidly, and each change may result in the need to re-design accessible interfaces. We are closer to goal of having assistive technology adaptations available when the mainstream consumer product ships, but there are still many problems with "workarounds" necessary to make mainstream operating system, productivity software and internet access accessible to people with disabilities.

Development and maintenance of access to ICT must be driven by the needs of people with disabilities. Developments which broaden the scope, applicability and usability of the human technology

interface will be driven, at least in part by the needs of people who have disabilities.

The Internet (e-mail and chat rooms) have the advantage of anonymity, and this can be a major benefit to individuals who have disabilities. Because the person's disability is not immediately visible, people who have disabilities report that they enjoy establishing relationships with people who experience them first as a person and then learn of their disability. For example, Blackstone, (1996) describes some of the advantages of e-mail for individuals who have disabilities. Since the receiver of the message reads it at a later time composition can be at a slower speed. The person with a disability can communicate with another person without someone else being present, establishing a greater sense of privacy than situations in which an attendant is required. It is also possible to work form any location-avoiding some transportation problems

3.3 Universal Design

Increasingly, commercial products are being designed to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design (NC State University, The Center for Universal Design, 1997).

3.3.1 General Principles of Universal Design

Features are built into products to make them more useful to persons who have disabilities (e.g., larger knobs; a variety of display options--visual, tactile, auditory; alternatives to reading text--icons, pictures) are built into the product. This is much less expensive than modifying a product after production to meet the needs of a person with a disability. The North Carolina State University Center for Universal Design, in conjunction with advocates of universal design, have compiled a set of principles of universal design, shown in Box 1. This center also maintains a Web site on universal design (www.design.ncsu.edu/cud).

3.3.2 Universal Design for ICT

In universal design for ICT the barriers are technological rather than political and economic barriers that characterize architectural and commercial product design (Emiliani, 2006). The goal of universal design for ICT is to have an environment with enough embedded intelligence to be easily adaptable. The features of future information services are that there will be no clearly predefined service and little distinction between interpersonal communication and access to information. Services will need to be highly interactive, inherently multimedia, sensory multimodal (i.e., access via auditory or visual means is equally possible). To achieve this cooperation between users or representatives of users is critical in a variety of contexts of use. The overall goal is to have access to information involving communities of users with a wide range of motor, sensory and cognitive skills.

ONE: EQUITABLE USE
The design is useful and marketable to people with diverse abilities.
TWO: FLEXIBILITY IN USE
The design accommodates a wide range of individual preferences and abilities.
THREE: SIMPLE AND INTUITIVE USE
Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level.
FOUR: PERCEPTIBLE INFORMATION
The design communicates necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities.
FIVE: TOLERANCE FOR ERROR
The design minimizes hazards and the adverse consequences of accidental or unintended actions.
SIX: LOW PHYSICAL EFFORT
The design can be used efficiently and comfortably and with a minimum of fatigue.
SEVEN: SIZE AND SPACE FOR APPROACH AND USE
Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility.

Box 1: Principles of Universal Design From North Carolina State University, The Center for. Universal Design, 1997.

In addition to Universal Design for ICT, access to capabilities of mainstream technologies includes individualized assistive technologies that are easily – customized. This in return requires an increased understanding of the biological/physical interface for the control of assistive technologies and expanded availability of embedded systems networks.

3.4 A Working Definition of Assistive Technologies

The *International Classification of Functioning, Disability and* Health (ICF) is a system developed by the World Health Organization (WHO) that is designed to describe and classify health and health related states. These two domains are described by

body factors (body structures and functions) and individual and societal elements (activities and participation) (WHO, 2001). The ICF recognizes two contextual factors that modify health and health related states: the environment and personal factors (WHO, 2001). Environmental elements include assistive technologies in relation to activities of daily living, mobility, communication, religion and spirituality as well as in specific contexts such as education, employment and culture, recreation and sport (WHO, 2001). Other environmental elements such as access to public and private buildings, and the natural and built outdoor environments, also have implications for assistive technologies.

A commonly used definition of assistive technology is from the Technical Assistance to the States Act in the United States (Public Law (PL) 100-407): ***Any item, piece of equipment or product system whether acquired commercially off the shelf, modified, or customized that is used to increase, maintain or improve functional capabilities of individuals with disabilities.***

3.4.1 Hard and Soft Technologies

Odor (1984) has distinguished between hard technologies and soft technologies. Hard technologies are readily available components that can be purchased and assembled into assistive technology systems. The main distinguishing feature of hard technologies is that they are tangible. On the other hand, soft technologies are the human areas of decision making, strategies, training, concept formation, and service delivery as described earlier in this chapter. Soft technologies are generally captured in one of three forms: (1) people, (2) written, and (3) computer (Bailey, 1997). These aspects of technology, without which the hard technology cannot be successful, are much harder to obtain. Soft technologies are difficult to acquire because they are highly dependent on human knowledge rather than tangible objects. This knowledge is obtained slowly through formal training, experience, and textbooks such as this one. The development of effective strategies of use also has a major effect on assistive technology system success. Initially the formulation of these strategies may rely heavily on the knowledge, experience, and ingenuity of the assistive technology practitioner. With growing experience, the assistive technology user originates strategies that facilitate successful device use. There is a false belief that progress is solely driven by "hard" technological change The gap between the general public and persons with disabilities can only be closed by gains in both soft and hard technologies

3.4.2 Mainstream Technologies to Specially Designed Technologies: A Range of Options

As illustrated in Figure 1, the needs of people with disabilities can be met in a number of ways. Off the shelf "standard" (i.e., mainstream technologies) commercially available devices (especially those designed using the principles of universal design) can often be used by people with a variety of disabilities. For example, standard personal computers designed for the general population are often used by persons with disabilities. Sometimes these need to be modified however, to make them useable. Another type of commercially available device is one that is mass-produced but specifically designed for individuals with disabilities *(special commercially available devices)*. These devices often need to be modified to meet the needs of a specific individual. Our goal is to reduce the amount of modification necessary and to make mainstream technologies as accessible as possible. However, there will always be a portion of the disabled population that will require specifically designed assistive technologies.

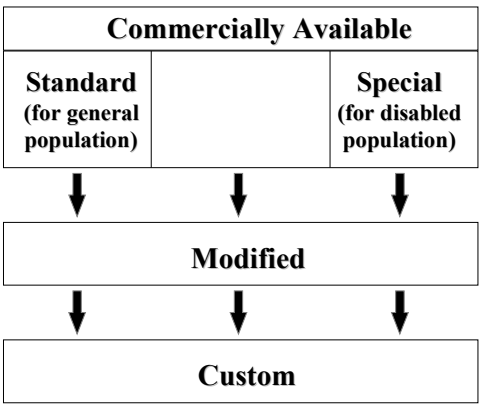


Figure 1: This diagram shows the progression from commercially available devices for the general population and commercially available devices for special populations to modified devices and custom devices. From Cook and Polgar, (2007).

3.5 The Human Technology Interface for ICT

3.5.1 General Concepts

It is estimated that as many as 40 million persons in the United States alone have physical, cognitive, or sensory disabilities (Lazzaro, 1999). The world-wide impact is significantly larger. If these people are to compete on an equal basis with non-disabled individuals, then it is extremely important that the internet be accessible to all. As the internet becomes more and more dependent on multimedia representations involving complex graphics, animation, and audible sources of information, the challenges for people who have disabilities increase. In order for access to the Internet to be useful to people with disabilities, the accessibility approach must be independent of individual devices. This means that users must be able to interact with a *user agent* (and the document it renders) using the input and output devices of their choice based on their specific needs. A **user agent** is defined as software to access Web content (www.w3.org/wai). This includes desktop graphical browsers, text and voice browsers, mobile phones, multimedia players, and software assistive technologies (e.g., screen readers, magnifiers) that are used with browsers. The person with a disability interacts with technology through the Human Technology Interface (HTI) (Cook and Polgar, 2007).

The graphical user interface (GUI) has both positive and negative implications for persons with disabilities. The positive features are those that apply to non-disabled users (e.g., use of icons, recognition rather than recall memory, screen icons for the same task look the same, operations such as opening and closing files are always the same). The GUI is the standard user interface because of its ease of operation for novices and its consistency of operation for experts. The latter ensures that every application behaves in basically the same way. People with motor disabilities may not have the necessary physical (eye-hand coordination) and visual skills to navigate the GUI. Modification of the GUI to allow specialized access (see Figure 1) can also be more challenging for GUI-based operating systems.

As networks are expanded and more devices (e.g., cell phones, PDAs) have open architectures, it will be possible to download profiles, adaptations and special instructions that enable adaptable systems to be developed to meet the needs of people who have disabilities. Some examples are (1) trainable hearing aids that adjust automatically to the environments in which they are used; (2) a "Smart House" that assesses occupants current state and the state of various home utilities to aid with common activities of daily living, provides feedback should residents become disoriented or confused and report medical emergencies automatically; an orientation and direction finding device that senses the current location (via GPS) and gives directions to a desired location for individuals who cannot read maps because of visual or cognitive disabilities.

3.5.2 Access for Motor Impairment

There are a significant number of people who cannot effectively use standard keyboards, mouse controls or switches. It is likely that we will see a much greater understanding of the biological/physical interface for the control of computers in the future (Applewhite, 2004).

One approach that may offer promise is the brain computer interface (BCI). BCI systems may be grouped into a set of functional components including the input device, amplification, feature extraction, feature translation and user feedback (Mason and Birch, 2003). Signals are mathematically analyzed to extract features useful for control (Fabiani, Mcfarland, Walla, and Pfurtscheller 2004). Features or signals that have been used include slow cortical potentials, P300 evoked potential, sensorimotor rhythms recorded from the cortex and neuronal action potentials recorded within the cortex). A typical task for a user is to visualize different movements or sensations or images.

Another approach to cursor control is the use of a digital camera and image recognition software to track a particular body feature to control an on-screen mouse cursor (Betke, Gips and Fleming, 2002). The most easily tracked feature is the tip of the nose, but the eye (gross eye position not point of gaze), lip, chin and thumb have also been used. Non-disabled subjects used this approach and fund that the camera mouse was accurate but slower than a typical hand-controlled mouse. Using an on-screen keyboard the camera mouse was half as fast as a regular mouse in a typing task, but the accuracy obtained was equivalent on each system. More and more computers have built-in cameras, so the camera mouse requires only software to capture the body feature image and interpret its movement as mouse commands. This may lead to wider application of this technique.

There are many other approaches that are used to provide access to and control over technologies for people with severe motor disabilities (Cook and Polgar, 2007) \. These range form keyboards of various type, to automatic speech recognition to mouse and mouse emulators systems to single and multiple switches.

3.5.3 Access for Cognitive Impairment

Cognitive disabilities include a wide range of skills and deficiencies. Learning disabilities typically involve significant difficulties in understanding or in using either spoken or written language, and these difficulties may be evident in problems with reading, writing, mathematical manipulation, listening, spelling or speaking (Edyburn, 2005). These limitations make it increasingly difficult to access complicated Web sites that may include flashing pictures, complicated charts, and large amounts of audio and video data. While there are assistive technologies that are specifically designed to address these areas (discussed later in this chapter), many of the technological tools are useful for all students, and are part of instructional technology (Ashton, 2005). Even the so-called assistive technologies have features (e.g., multimedia, synthetic speech output, voice recognition input) that are useful to all learners.

For individuals with acquired cognitive disabilities due to injury (e.g., traumatic brain injury) or disease (e.g., stroke (CVA) or dementia) changing features such as font size, background/ foreground color combinations, contrast, spacing between words, letters and paragraphs and using graphics can all improve access to screen-based information. Another technological concept for these individuals is a cognitive prosthesis, which is a custom-designed computer-based compensatory strategy that directly assists in performing daily activities[1]. It may also include additional technologies such as a cell phone, pager, digital camera or low tech approaches

Persons with intellectual disabilities have difficulties with memory, language use and communication, abstract conceptualization, generalization and problem identification/problem solving. Characteristics of the HTI that are important for these individuals include simplicity of operation, capacity of the technology to support repetition, consistency in presentation, and inclusion of multiple modalities (e.g., speech, sounds and graphical representations) (Wehmeyer, Smith and Davies, 2005).

An example of technology designed for cognitive needs is the Planning and Execution Assistant and Trainer (PEAT). It is a PDA-based personal planning assistant designed to assist individuals with cognitive disorders due to brain injury, stroke, Alzheimer's disease, and similar conditions (Levinson, 1997). PEAT employs artificial intelligence to automatically generate plans and also to revise those plans when unexpected events occur. PEAT uses a combination of manually entered schedules and a library of stored scripts describing activities of daily living (e.g., morning routine or shopping). Scripts can be used for both planning and for execution. Planning involves a simulation of the activity with key decision points presented and prompts (auditory and visual) supplied necessary to aid the individual through the planning process. The plan to be executed can be either the stored script or a modified script based on the simulation. The PEAT artificial intelligence software generates the best strategy to execute the required steps in the plan (LoPresti, Mihailidis, and Kirsch, 2004). PEAT also automatically monitors performance, and corrects schedule problems when necessary.

3.5.4 Access for Auditory Impairment

Since web pages are a mixture of text, graphics, and sound, people who are deaf may be prevented from accessing some information unless alternative methods are available. The primary approach for thee individual is the use of the Microsoft Synchronized Accessible Media Interchange (SAMI), which allows authors of Web pages and multimedia software to add closed captioning for users who are deaf or hard of hearing. This approach is similar to the use of closed captioning for television viewers. The W3C WAI SMIL (www.w3.org/WAI) is designed to facilitate multimedia presentations in which an author can describe the behavior of a multimedia presentation, associate hyperlinks with media objects, and describe the layout of the presentation on a screen

Trainable hearing aids adjust automatically to the environments in which they are used through access to embedded information networks. This allows automatic adaptation to changing noise levels and environments.

[1] Institute for Cognitive Prosthetics, http://www.brain-rehab.com/definecp.htm

3.5.5 Access for Visual Impairment

The W3C WAI user agent guidelines are based on several principles that are intended to improve the design of both types of user agents. The first is to ensure that the user interface is accessible. This means that the consumer using an adapted input system must have access to the functionality offered by the user agent through its user interface. Second, the user must have access to document content through the provision of control of the style (e.g., colors, fonts, speech rate, and speech volume) and format of a document. A third principle is that the user agent help orient the user to where he is in the document or series of documents. In addition to providing alternative representations of location in a document (e.g., how many links the document contains or the number of the current link), a well-designed navigation system that uses numerical position information allows the user to jump to a specific link. Finally, the guidelines call for the user agent to be designed following system standards and conventions. These are changing rapidly as development tools are improved.

Communication through standard interfaces is particularly important for graphical desktop user agents, which must make information available to assistive technologies. Technologies such as those produced by the W3C include built-in accessibility features that facilitate interoperability. The standards being developed by the W3C WAI provide guidance for the design of user agents that are consistent with these principles. The guidelines are available on the W3C WAI Web page (www.w3.org/wai).

3.5.6 Other ICT Access

Cellular telephones are becoming more powerful with capabilities approaching that of personal computers. This expanded capability will provide significant advantages for people with disabilities, especially those with low vision or blindness. describes Three changes will be particularly valuable to people who have disabilities:: (1) standard cell phones will have sufficient processing power for almost all the requirement of persons with visual impairments, (2) software will be able to be downloaded into these phones easily, (3) wireless connection to a worldwide network will provide a wide range of information and services in a highly mobile way (Fruchterman, 2003). Because many of these features will be built into standard cell phones the cost will be low and reachable by persons with disabilities. A major advance will occur if the cell phone industry moves away from proprietary software to an open source format providing the basis for a greater diversity of software for tasks such as text-to-speech output, voice recognition and optical character recognition in a variety of languages. Many applications for people with disabilities will be able to be downloaded from the internet. With expanded availability of embedded systems, it will be possible for a user to store their customized programs on the network and download them as needed form any remote location.

Downloading a talking book program into a cell phone can provide access to digital libraries for persons who are blind. Outputs in speech or enlarged visual displays can be added as needed by the user. With a built-in camera and network access a blind person could obtain a verbal description of a scene by linking to on-line volunteers who provide descriptions of images. These applications will depend on the increasing application of universal design in information technology products (Tobias, 2003). These applications include ATMs, cell phones, vending machines and other systems that are encountered on a daily basis (Tobias, 2003).

4 INFRASTRUCTURE FOR FUTURE ACCESSIBILITY

The infrastructure for future accessibility consists of: (1) an expanded, smarter and more available "real" and "virtual" internet, (2) Home automation systems that are smarter and have greater interconnectivity, (3) universal design principles that are applied more widely, (4) alternative approaches for accessing information technologies, and (5) special-purpose assistive technologies.

The Infrastructure for future accessibility will depend on several factors. These include: Web-based virtual systems, home automation, universal design for ICT, alternatives for accessing information technologies and special-purpose assistive technologies. In addition there is n on going need for the development of soft technology tools.

If ICT advances are not adaptable enough to be accessible to persons with disabilities it will further increase the disparity between those individuals and the rest of the population leading to further isolation and economic disadvantage. On the other hand, availability of these technologies in a transparent way will contribute to full inclusion of individuals who have disabilities in the mainstream of society.

5 CONCLUSIONS

The move to the information age offers great promise for persons with disabilities. It also holds great threats for persons with disabilities. Constant vigilance is required to insure that information technologies remain accessible and responsive to the needs of persons with disabilities. The future for persons with disabilities will not be driven by advances in technology, but rather by how well we can take advantage of those advances for the accomplishment of the many tasks of living that require technological assistance

6 SUMMARY

Anticipated changes in technologies coupled with the focus on the social aspects of disability, provide a significant opportunity for major advances in the degree to which individuals with disabilities can participate in all aspects of life, including work, school, leisure and self care.

Technological advances will be particularly important as the percentage of the population that is elderly rises. Concepts from universal design will be important in ensuring that this segment of the population remains active and is able to participate in society. This new group of elderly individuals will also be more experienced with computers and other technologies than their predecessors and they may well demand greater performance and adaptability from both assistive technologies and mainstream ICT (e.g., telephones, internet communication).

The percentage of individuals with long-term disabilities who join the over 65 age group will also increase. These individuals will have been long-term users of assistive technologies, and their experience will have major implications for developments to meet future needs.

While much of what I have described is conjecture, it is based on modest extrapolation from the current state of the art. There are some things that we know with a high degree of certainty. We know that computer systems will be faster, have more memory be smaller and be less expensive for the same or greater functionality. We also know that the communication channel bandwidth will continue to increase allowing much more information and much more sophisticated information processing. Finally, it is clear that people with disabilities will continue to assert their right to fully participate in society.

Technological advances also raise questions for people who have disabilities. The most important of these is whether accessibility will keep pace with technological developments. For example, will assistive technologies for input and output be compatible with the user agents and operating systems of tomorrow. A second major question is whether the needs of persons with disabilities will be a driving force in future technological developments. Will people who have disabilities have to adapt to the existing technologies based on characteristics for non-disabled people or will universal design become a greater reality? In the latter case, adaptations will become less important and accessibility will become the rule rather than the exception.

For people who have disabilities, there are significant implications of emerging information processing technologies. If not closely monitored, these could result in less rather than more access to the new information economy for persons with disabilities. Despite the wider use of universal design principles, there will still be a need for effective assistive technology design and application if individuals with disabilities are to realize the full potential of the new information age.

REFERENCES

Applewhite A: (2004). 40 years: the luminaries. *IEEE Spectrum,* 41(11), 37-58.

Ashton T.M. (2005). Students with learning disabilities using assistive technology in the inclusive classroom, pp. 229-238, in Edyburn D., Higgins K.. & Boone R. (eds.): *Handbook of Special Education Technology Research and Practice,* Whitefish Bay, Wisconsin: Knowledge by Design, Inc.

Bailey, R.W: (1996). *Human performance engineerin.,* ed 2, Upper Saddle River, NJ, , Prentice Hall.

Betke M., Gips J. & Fleming P: (2002). The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Trans. Neural Systems and Rehabilitation Engineering,* 10(1): 1-10.

Blackstone S. (1996). The Internet: what's the big deal *Augment Commun News* 9(4):1-5.

Cook A.M. & Polgar J.M. (2007). *Cook and Husey's Assistive Technologies: Principles and Practice*, 3rd ed St. Louis: Elsevier.

Edyburn D.L. (2005). Assistive technology and students with mild disabilities: from consideration to outcome measurement, pp 239-270, in Edyburn D, Higgins K. & Boone R. (eds.): *Handbook of Special Education Technology Research and Practice,* Whitefish Bay, Wisconsin: Knowledge by Design, Inc.,

Emiliani, P.L. (2006). Assistive technology (AT) versus Mainstream Technology (MST): The research perspective. *Tech Disab*, 18, 19-29,

Fabiani G.E., Mcfarland D.J., Wolpaw J.R., & Pfurtscheller G (2004). Conversion of EEG activity into cursor movement by a brain-computer interface (BCI). *IEEE Trans. Neural Systems and Rehab Engr*, 12: 331-338.

Fruchterman J.R: (2003) In the palm of your hand: A vision of the future of technology for people with visual impairments. *J Vis Impair Blindness, 97*(10), 585-591.

Kumagai, J. (2004).Talk to the machine. *IEEE Spectrum,* 39(9): 60- 64.

Lazzaro J.L. (1999). Helping the web help the disabled, *IEEE Spectrum* 36(3): 54-59

Levinson R.L. (1997): The planning and execution assistant and trainer. *J. head trauma rehabilitation*, 12(2): 769-775.

Lewis, J.R. (2004). .In the eye of the beholder*, IEEE Spectrum,* 41(5)24- 28.

LoPresti E.F., Mihailidis A. & Kirsch N. (2004): Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation,* 14(1): 5-39.

Mason S.G. & Birch G.E: (2003).A general framework for brain-computer interface design. *IEEE Trans. Neural Systems and Rehab Engr*, 11, 70-85.

Odor P. (1984): Hard and soft technology for education and communication for disabled people, *Proc Int Comp Conf*, Perth, Western Australia,

Tobias J: (2003). Information technology and universal design: An agenda for accessible technology. *J Vis Impair Blindness, 97*(10), 592-601.

Ungson G.R. and Trudel J.D. (1999). The emerging knowledge-based economy. *IEEE Spectrum*, 36(5):60-65.

Wehmeyer M.L., Smith S.J., Palmer SB, Davies D.K. & Stock S.E: (2004). Technology use and people with mental retardation. *International Review of Research in Mental Retardation,* 29: 291-337.

World Health Organization (2001): *International classification of functioning disability and health*-ICF, Geneva, World Health Organization

Wright, RA. (2004). *A Short History of Progress*, Toronto: Anansi Publications.

FULL PAPERS

EARS: ELECTROMYOGRAPHICAL AUTOMATIC RECOGNITION OF SPEECH

Szu-Chen Stan Jou and Tanja Schultz
International Center for Advanced Communication Technologies
Carnegie Mellon University, Pittsburgh, PA, USA
Karlsruhe University, Karlsruhe, Germany
scjou@cs.cmu.edu, tanja@cs.cmu.edu

Keywords: Electromyography, Speech Recognition, Articulatory Feature, Feature Extraction.

Abstract: In this paper, we present our research on automatic speech recognition of surface electromyographic signals that are generated by the human articulatory muscles. With parallel recorded audible speech and electromyographic signals, experiments are conducted to show the anticipatory behavior of electromyographic signals with respect to speech signals. Additionally, we demonstrate how to develop phone-based speech recognizers with carefully designed electromyographic feature extraction methods. We show that articulatory feature (AF) classifiers can also benefit from the novel feature, which improve the F-score of the AF classifiers from 0.467 to 0.686. With a stream architecture, the AF classifiers are then integrated into the decoding framework. Overall, the word error rate improves from 86.8% to 29.9% on a 100 word vocabulary recognition task.

1 INTRODUCTION

As computer technologies advance, computers have become an integral part of modern daily lives and our expectations for a user-friendly interface grow everyday. Automatic speech recognition (ASR) is one of the most efficient front-end for human-computer interface because it is natural for humans to communicate through speech. ASR is an automatic computerized speech-to-text process which converts human speech signals into written words. It has various applications, such as voice command and control, dictation, dialog systems, audio indexing, speech-to-speech translation, etc. However, these ASR applications usually do not work well in noisy environments. Besides, they usually require the user to speak aloud, which may be disturbing to bystanders and brings up concern of privacy loss. In this paper, we describe our research of integrating signals based on electromyography with traditional acoustic speech signals for the purpose of speech recognition.

The input speech signal of the traditional ASR process is usually recorded with a microphone, e.g., a close-talking headset or a telephone. However, from the ASR point of view, microphone recordings often suffer from ambient noise or in other words the noise robustness issue, because microphones pick up vibration from the air-transmitted channel; therefore, while picking up air vibration generated by human voices, microphones also pick up air-transmitted ambient noises. In most cases, ambient noise deteriorates the ASR performance and the decrease in performance depends on how badly the original voice signal has been corrupted by noise. Besides the noise robustness issue, microphone-based ASR often has applicability issues, by which we mean that it is often suboptimal to use microphones as the input device of speech applications in certain situation. For example, in an on-line shopping system, it is often required to input confidential information such as credit card numbers, which may be overheard if the user speak aloud via the air-transmitted channels. Usually this kind of overhearing results in confidentiality or privacy infringement. Besides, another issue of applicability is that speaking aloud usually annoys other people. Just imagine how annoying it would be if your officemate spends all day dictating to the computer to write a report, let alone many people dictate simultaneously.

In order to resolve the noise robustness and the applicability issues, we have applied electromyographic (EMG) method to our speech recognition research. The motivation is that the EMG method is inherently robust to ambient noise and it enables silent speech recognition to avoid disturbance and confidentiality issues. The EMG method measures muscular electric potential with a set of electrodes attached to the skin where the articulatory muscles underlie. In the physi-

ological speech production process, as we speak, neural control signals are transmitted to articulatory muscles, and the articulatory muscles contract and relax accordingly to produce voice. The muscle activity alters the electric potential along the muscle fibers, and the EMG method can measure this kind of potential change. In other words, the articulatory muscle activities result in electric potential change, which can be picked up by EMG electrodes for further signal processing, e.g., speech recognition. The EMG method is inherently robust to ambient noise because the EMG electrodes contact to the human tissue directly without the air-transmission channel. In addition, the EMG method has better applicability because the EMG method makes it possible to recognize silent speech, which means mouthing words without making any sound.

For silent speech recognition with EMG, Manabe et al. first showed that it is possible to recognize five Japanese vowels and ten Japanese isolated digits using surface EMG signals recorded with electrodes pressed on the facial skin (Manabe et al., 2003; Manabe and Zhang, 2004). EMG has been a useful analytic tool in speech research since the 1960's (Fromkin and Ladefoged, 1966), and the recent application of surface EMG signals to automatic speech recognition was proposed by Chan et al. They focused on recognizing voice command from jet pilots under noisy environment, so they showed digit recognition in normal audible speech (Chan et al., 2002). Jorgensen et al. proposed sub auditory speech recognition using two pairs of EMG electrodes attached to the throat. Sub vocal isolated word recognition was demonstrated with various feature extraction and classification methods (Jorgensen et al., 2003; Jorgensen and Binsted, 2005; Betts and Jorgensen, 2006). Maier-Hein et al. reported non-audible EMG speech recognition focusing on speaker and session independency issues. (Maier-Hein et al., 2005).

However, these pioneering studies are limited to small vocabulary ranging from five to around forty isolated words. The main reason of this limitation is that the classification unit is restrained to a whole utterance, instead of a phone as a smaller and more flexible unit. As a standard practice of large vocabulary continuous speech recognition (LVCSR), the phone is a natural unit based on linguistic knowledge. From the pattern recognition's point of view, the phone as a smaller unit is preferred over a whole utterance because phones get more training data per classification unit for more reliable statistical inference. The phone unit is also more flexible in order to constitute any pronunciation combination of words as theoretically unlimited vocabulary for speech recognition. With the phone unit relaxation, EMG speech recognition can be treated as a standard LVCSR task and we can apply any advanced LVCSR algorithms to improve the EMG speech recognizer.

In this paper, we introduce such an EMG speech recognition system with the following research aspects. Firstly, we analyze the phone-based EMG speech recognition system with articulatory features and their relationship with signals of different EMG channels. Next, we demonstrate the challenges of EMG signal processing with the aspect of feature extraction for the speech recognition system. We then describe our novel EMG feature extraction methods which makes the phone-based system possible. Lastly, we integrate the novel EMG feature extraction methods and the articulatory feature classifiers into the phone-based EMG speech recognition system with a stream architecture. Notice that the experiments described in this paper are conducted on normal audible speech, not silent mouthing speech.

2 RESEARCH APPROACH

2.1 Data Acquisition

In this paper, we report results of data collected from one male speaker in one recording session, which means the EMG electrode positions were stable and consistent during this whole session. In a quiet room, the speaker read English sentences in normal audible speech, which was simultaneously recorded with a parallel setup of an EMG recorder and a USB soundcard with a standard close-talking microphone attached to it. When the speaker pressed the push-to-record button, the recording software started to record both EMG and speech channels and generated a marker signal fed into both the EMG recorder and the USB soundcard. The marker signal was then used for synchronizing the EMG and the speech signals. The speaker read 10 times of a set of 38 phonetically-balanced sentences and 10 times of 12 sentences from news articles. The 380 phonetically-balanced utterances were used for training and the 120 news article utterances were used for testing. The total duration of the training and test set are 45.9 and 10.6 minutes, respectively. We also recorded ten special silence utterances, each of which is about five seconds long on average. The format of the speech recordings is 16 kHz sampling rate, two bytes per sample, and linear PCM, while the EMG recording format is 600 Hz sampling rate, two bytes per sample, and linear PCM. The speech was recorded with a Sennheiser HMD 410 close-talking headset.

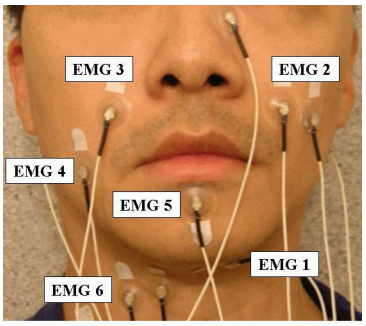


Figure 1: EMG positioning.

The EMG signals were recorded with six pairs of Ag/Ag-Cl surface electrodes attached to the skin, as shown in Fig. 1. Additionally, a common ground reference for the EMG signals is connected via a self-adhesive button electrode placed on the left wrist. The six electrode pairs are positioned in order to pick up the signals of corresponding articulatory muscles: the levator angulis oris (EMG2,3), the zygomaticus major (EMG2,3), the platysma (EMG4), the orbicularis oris (EMG5), the anterior belly of the digastric (EMG1), and the tongue (EMG1,6) (Chan et al., 2002; Maier-Hein et al., 2005). Two of these six channels (EMG2,6) are positioned with a classical bipolar configuration, where a 2cm center-to-center inter-electrode spacing is applied. For the other four channels, one of the electrodes is placed directly on the articulatory muscles while the other electrode is used as a reference attached to either the nose (EMG1) or to both ears (EMG 3,4,5).

In order to reduce the impedance at the electrode-skin junctions, a small amount of electrode gel was applied to each electrode. All the electrode pairs were connected to the EMG recorder (Becker, 2005), in which each of the detection electrode pairs pick up the EMG signal and the ground electrode provides a common reference. EMG responses were differentially amplified, filtered by a 300 Hz low-pass and a 1Hz high-pass filter and sampled at 600 Hz. In order to avoid loss of relevant information contained in the signals we did not apply a 50 Hz notch filter which can be used for the removal of line interference. Also note that all care was taken such that wearing the close-talking headset does not interfere with the EMG electrode attachment.

2.2 EMG-based Speech Recognition

We used the following approach to bootstrap the phone-based EMG speech recognizer. First of all, the forced alignment of the audible speech data is generated with a Broadcast News (BN) speech recognizer (Yu and Waibel, 2000), which is trained with the Janus Recognition Toolkit (JRTk). Since we have parallel recorded audible and EMG speech data, the forced-aligned labels of the audible speech were used to bootstrap the EMG speech recognizer. Since the training set is very small, we only trained context-independent acoustic models. The trained acoustic model was used together with a trigram BN language model for decoding. Because the problem of large vocabulary continuous speech recognition is still very difficult for state-of-the-art EMG speech processing, we restricted the decoding vocabulary to the words appearing in the test set in this study. This approach allows us to better demonstrate the performance differences introduced by different feature extraction methods. To cover all the test sentences, the decoding vocabulary contains 108 words in total. Note that the training vocabulary contains 415 words, 35 of which also exist in the decoding vocabulary. Also note that the test sentences were not applied for language model training.

2.3 Articulatory Feature Classifier and Stream Architecture

Compared to widely-used cepstral features for automatic speech recognition, articulatory features are expected to be more robust because they represent articulatory movements, which are less affected by speech signal variation or noise. Instead of measuring the AFs directly, we derive them from phones as described in (Metze and Waibel, 2002). More precisely, we use the IPA phonological features for AF derivation. In this work, we use AFs that have binary values. For example, each of the positions of the dorsum, namely FRONT, CENTRAL and BACK is an AF that has a value either present or absent. To classify the AF as present or absent, the likelihood scores of the corresponding present model and absent model are compared. Also, the models take into account a prior value based on the frequency of features in the training data.

The training of AF classifiers is done on middle frames of the phones only, because they are more stable than the beginning or ending frames. Identical to the training of EMG speech recognizer, the AF classifiers are also trained solely on the EMG signals without speech acoustics. There are 29 AF classifiers, each of which is a Gaussian Mixture Model (GMM) containing 60 Gaussians. To test the performance, the AF classifiers are applied and generate frame-based hypotheses.

The idea behind the stream architecture with AF classifiers is that the AF streams are expected to provide additional robust phonological information to the

phone-based hidden Markov model (HMM) speech recognizer. The stream architecture employs a list of parallel feature streams, each of which contains one of the acoustic or articulatory features. Information from all streams are combined with a weighting scheme to generate the EMG acoustic model scores for decoding (Metze and Waibel, 2002).

2.4 Feature Extraction

2.4.1 Traditional Spectral Feature

The recorded EMG signal is transformed into 18-dimensional feature vectors, with 54-ms observation window and 10-ms frame-shift for each channel.

For each channel, hamming-windowed Short Time Fourier Transform is computed, and then its delta coefficients serve as the first 17 coefficients of the final feature. The 18th coefficient consists of the mean of the time domain values in the given observation window (Maier-Hein et al., 2005). In the following experiments, features of one or more channels can be applied. If more than one channel are used for classification, the features of the corresponding channels are concatenated to form the final feature vector.

2.4.2 Special EMG Feature

Since the EMG signal is very different from the speech signal, it is necessary to explore feature extraction methods that are suitable for EMG speech recognition. Here we describe the signal preprocessing steps and feature extraction methods we designed for EMG signals.

As noted above, the EMG signals vary across different sessions. Nonetheless, the DC offsets of the EMG signals vary, too. In the attempt to make the DC offset zero, we estimate the DC offset from the special silence utterances on a per session basis, then all the EMG signals are preprocessed to subtract this session-based DC offset. Although we only discuss a single session of a single speaker in this paper, we expect this DC offset preprocessing step makes the EMG signals more stable.

To describe the features designed for EMG signals, we denote the EMG signal with normalized DC as $x[n]$ and its short-time Fourier spectrum as $\mathbf{X}$. We also denote the nine-point double-averaged signal $w[n]$, high frequency signal $p[n]$, and the corresponding rectified signal $r[n]$.

We then define the time-domain mean features $\bar{\mathbf{x}}, \bar{\mathbf{w}}$, and $\bar{\mathbf{r}}$ of the signals $x[n], w[n]$, and $r[n]$, respectively. Besides, we use the power features $\mathbf{P_w}$ and $\mathbf{P_r}$ and we define $\mathbf{z}$ as the frame-based zero-crossing rate of $p[n]$.

To better model the context, we use the following contextual filters, which can be applied on any feature to generate a new one. The delta filter: $D(\mathbf{f}_j) = \mathbf{f}_j - \mathbf{f}_{j-1}$. The trend filter: $T(\mathbf{f}_j, k) = \mathbf{f}_{j+k} - \mathbf{f}_{j-k}$. The stacking filter: $S(\mathbf{f}_j, k) = [\mathbf{f}_{j-k}, \mathbf{f}_{j-k+1}, ..., \mathbf{f}_{j+k-1}, \mathbf{f}_{j+k}]$, where j is the frame index and k is the context width. Note that we always apply linear discriminant analysis (LDA) on the final feature in order to reduce the dimensionality to 32.

3 EXPERIMENTS AND ANALYSES

The performance metrics used in this paper are F-score and word error rate (WER). F-score ($\alpha = 0.5$) is reported for the AF performances and WER is reported for the speech recognition performances[1].

3.1 Articulatory Feature Analysis

3.1.1 Baseline System

First of all, we forced-aligned the speech data using the aforementioned BN system. In the baseline system, this time-alignment was used for both the speech and the EMG signals. Because we have a marker channel in each signal, the marker signal is used to offset the two signals to get accurate time-synchronization. Then the aforementioned AF training and testing procedures were applied both on the speech and the six-channel concatenated EMG signals. The averaged F-scores of all 29 AFs are 0.814 for the speech signal and 0.467 for the EMG signal. Fig. 2 shows individual AF performances for the speech and EMG signals along with the amount of training data. We can see that the amount of training data (given in frames of 10 ms) has an impact on the EMG AF performance.

3.1.2 Channel Synchronization

It is observed that human articulator movements are anticipatory to the speech signal as speech signal is a product of articulator movements and source excitation (Chan et al., 2002). This means the time alignment we used for bootstrapping our EMG-based

[1] With $\alpha = 0.5$, F-score $= 2PR/(P+R)$, where precision $P = C_{tp}/(C_{tp}+C_{fp})$, recall $R = C_{tp}/(C_{tp}+C_{fn})$, C_{tp} = true positive count, C_{fp} = false positive count, C_{fn} = false negative count.

WER $= \frac{S+D+I}{N}$, where S = word substitution count, D = word deletion count, I = word insertion count, N = number of reference words.

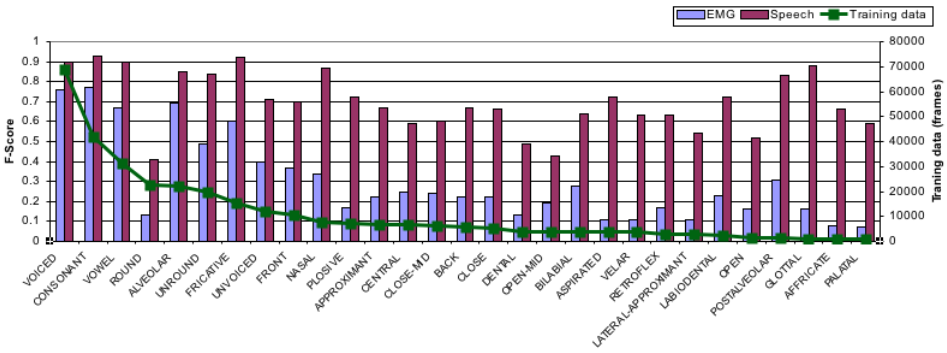


Figure 2: Baseline F-scores of the EMG and speech signals vs. the amount of training data.

system is actually mis-aligned for the EMG signals, because the speech and the EMG signals are inherently asynchronous in time. Based on this, we delayed the EMG signal with various duration to the forced-alignment labels of speech signal, and conducted the training and testing experiments respectively. As shown in Fig. 3, the initial time-alignment does not have the best F-score, while the best F-scores come with time delays around 0.02 second to 0.12 second. This result suggests that a time-delayed effect exists between the speech and the EMG signals.

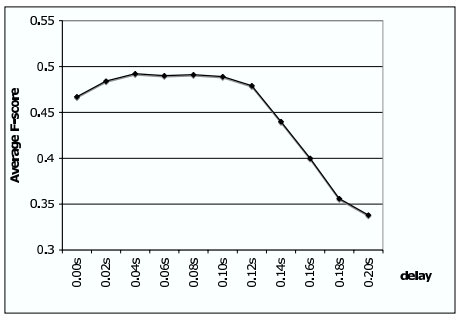


Figure 3: F-scores of concatenated six-channel EMG signals with various time delays (a delay of 0.1 means that the EMG signal is delayed to the acoustic signal by 0.1 seconds).

3.1.3 Articulator-Dependent Synchronization

To explore the time-delayed effect of EMG signals, we conducted the same experiments on the level of single EMG channels, instead of previously concatenated six-channels. The rationale is that articulators' behaviors are different from each other, so the resulted time delays are different on the corresponding EMG signals. The effect of different time delays can be seen in Fig. 4. We observed that some EMG signals are more sensitive to time delay than others, e.g. EMG1 vs. EMG6, where EMG6 is more consistent with different time delays. The delays to achieve peak performance vary for each channel and the variation is within the range of 0.02 to 0.10 seconds. To further show the time-delay effect, we also conducted an experiment which is identical to the baseline, except each channel is offset with its known best time delay. This approach gave a better F-score of 0.502 than the baseline's 0.467. It also outperforms the uniform delay of 0.04 second which gave 0.492.

Figure 4: F-scores of single-channel EMG signals with various time delays with respect to the speech signals.

3.1.4 Complementary EMG Pairs

As suggested in (Maier-Hein et al., 2005), concatenated multi-channel EMG features usually work better than single-channel EMG features. Therefore, based on aforementioned time-delayed results, we conducted experiments on EMG-pairs in which each EMG signal is adjusted with its best single-channel time offset. The first row of values in Table 1 shows the F-scores of single-channel baseline (i.e. without any time delay) and the second row shows those with the best single-channel time delay, while the rest of the values are F-scores of EMG pairs. The F-scores suggest that some EMG signals are complementary to each other, e.g. EMG1-3 and EMG2-6, which pairs perform better than both their single channels do.

Table 1: F-Score of EMG and EMG Pairs.

F-Scores	EMG1	EMG2	EMG3	EMG4	EMG5	EMG6
single	0.435	0.399	0.413	0.404	0.357	0.440
+delay	**0.463**	0.419	0.435	0.415	0.366	0.450
EMG1		0.439	**0.465**	0.443	0.417	0.458
EMG2			0.440	0.443	0.414	**0.464**
EMG3				0.421	0.414	0.449
EMG4					0.400	0.433
EMG5						0.399

3.1.5 Performance of Individual Articulators

In Table 2 and 3, we list the top-4 articulators that have the best F-scores. For single channels, EMG1 performs the best across these top-perfomance articulators, while EMG1-3, EMG1-6, and EMG2-6 are the better paired channels. Interestingly, even though EMG5 performs the worst as a single channel classifier, EMG5 can be complemented with EMG2 to form a better pair for VOWEL. In Fig. 5, we show six AFs that represent different characteristics of performance changes with different delays. For example, VOICED's F-scores are rather stable with various delay values while BILABIAL is rather sensitive. However, we do not have conclusive explanation on the relation between the AFs and the delays. Further exploration shall be conducted.

Table 2: Best F-Scores of Single EMG channels w.r.t. AF.

AFs	VOICED	CONSONANT	ALVEOLAR	VOWEL
	1 0.80	2 0.73	1 0.65	1 0.59
Sorted	6 0.79	3 0.72	3 0.61	2 0.59
F-score	3 0.76	1 0.71	2 0.59	6 0.56
	4 0.75	6 0.71	6 0.56	3 0.52
	2 0.74	4 0.69	4 0.55	4 0.51
	5 0.74	5 0.63	5 0.45	5 0.51

Table 3: Best F-Scores of Paired EMG Channels w.r.t. AF

AFs	VOICED	CONSONANT	ALVEOLAR	VOWEL
	1-6 0.77	1-6 0.76	1-3 0.69	2-6 0.64
Sorted	1-3 0.76	2-3 0.75	1-6 0.67	2-4 0.62
F-Score	1-2 0.76	3-6 0.74	1-2 0.66	2-5 0.62
	2-6 0.75	2-4 0.74	2-6 0.66	1-6 0.62
	3-6 0.75	2-6 0.74	2-3 0.65	1-3 0.61

3.2 Feature Extraction Experiments

In the following experiments, the final EMG features are generated by stacking single-channel EMG features of channels 1, 2, 3, 4, 6. We do not use channel 5 because it is relatively noisy for this experiment. The final LDA dimensions are reduced to 32 for all the experiments.

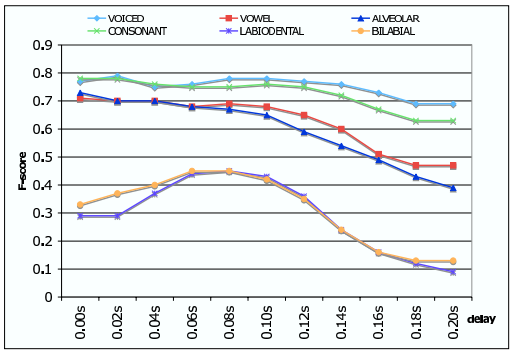


Figure 5: Performances of six representative AFs.

3.2.1 EMG ASR using Spectral Features

It was reported that the spectral coefficients are better than cepstral and LPC coefficients on EMG speech recognition (Maier-Hein et al., 2005). Therefore, we use the spectral features as baseline in this paper. As their WER is shown in Fig. 6, the spectral features are $\mathbf{S0} = \mathbf{X}$, $\mathbf{SD} = [\mathbf{X}, D(\mathbf{X})]$, and $\mathbf{SS} = S(\mathbf{X}, 1)$. We can see that the contextual features improve WER. Additionally, adding time delays for modeling the anticipatory effects also helps. This is consistent to the articulatory feature analysis above.

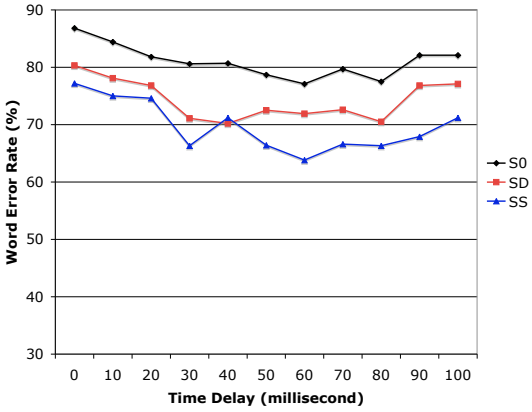


Figure 6: Word Error Rate on Spectral Features.

3.2.2 EMG ASR Systems using Spectral+temporal (ST) Features

Is was also reported that the time-domain mean feature provided additional gain to spectral features (Maier-Hein et al., 2005). Here we also added the time-domain mean feature, as their WER is shown in Fig. 7: $\mathbf{S0M} = \mathbf{X_m}$, $\mathbf{SDM} = [\mathbf{X_m}, D(\mathbf{X_m})]$, $\mathbf{SSM} = S(\mathbf{X_m}, 1)$, and $\mathbf{SSMR} = S(\mathbf{X_{mr}}, 1)$. where $\mathbf{X_m} = [\mathbf{X}, \bar{\mathbf{x}}]$ and $\mathbf{X_{mr}} = [\mathbf{X}, \bar{\mathbf{x}}, \bar{\mathbf{r}}, \mathbf{z}]$.

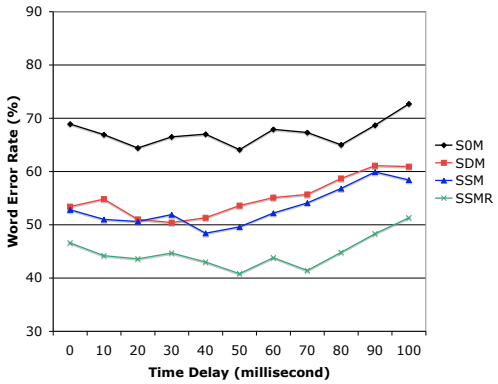


Figure 7: Word Error Rate on Spectral+Temporal Features.

3.2.3 EMG ASR Systems using EMG Features

We have observed that even though the spectral features are among the better ones, they are still very noisy for acoustic model training. Therefore we designed the EMG features that are normalized and smoothed in order to extract features from EMG signals in a more robust fashion. The performance of the EMG features are shown in Fig. 8, where the EMG features are

$$\mathbf{E0} = [\mathbf{f0}, D(\mathbf{f0}), D(D(\mathbf{f0})), T(\mathbf{f0}, 3)],$$
$$\text{where } \mathbf{f0} = [\bar{\mathbf{w}}, \mathbf{P_w}]$$
$$\mathbf{E1} = [\mathbf{f1}, D(\mathbf{f1}), T(\mathbf{f1}, 3)],$$
$$\text{where } \mathbf{f1} = [\bar{\mathbf{w}}, \mathbf{P_w}, \mathbf{P_r}, \mathbf{z}]$$
$$\mathbf{E2} = [\mathbf{f2}, D(\mathbf{f2}), T(\mathbf{f2}, 3)],$$
$$\text{where } \mathbf{f2} = [\bar{\mathbf{w}}, \mathbf{P_w}, \mathbf{P_r}, \mathbf{z}, \bar{\mathbf{r}}]$$
$$\mathbf{E3} = S(\mathbf{E2}, 1)$$
$$\mathbf{E4} = S(\mathbf{f2}, 5)$$

The essence of the design of feature extraction methods is to reduce noise while keeping the useful information for classification. Since the EMG spectral feature is noisy, we decide to first extract the time-domain mean feature, which is empirically known to be useful in literature. By adding power and contextual information to the time-domain mean, **E0** is generated and it already outperforms all the spectral-only features. Since the mean and power only represent the low-frequency components, we add the high-frequency power and the high-frequency zero-crossing rate to form **E1**, which gives us another 10% improvement. With one more feature of the high-frequency mean, **E2** is generated. **E2** again improves the WER. **E1** and **E2** show that the specific high-frequency information can be helpful. **E3** and **E4** use different approaches to model the contextual information, and they show that large context provides useful information for the LDA feature optimization step. They also show that the features with large context are more robust against the EMG anticipatory effect. We summarize by showing the performance of all the presented feature extraction methods in Fig. 9, in which all the feature extraction methods apply a 50-ms delay.

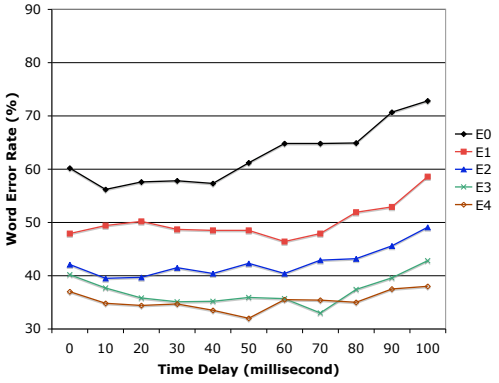


Figure 8: Word Error Rate on EMG Features.

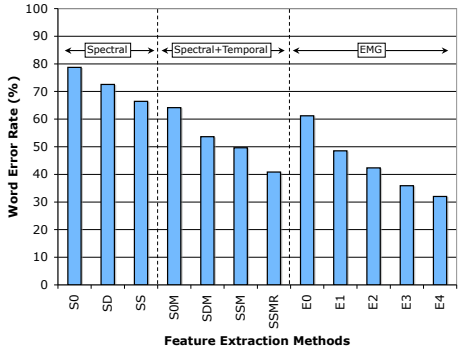


Figure 9: WER of Feature Extraction Methods.

3.3 Integration of Special EMG Feature and AF Classifiers

3.3.1 AF Classification with the E4 Feature

Identical to the aforementioned experiments, we forced-aligned the speech data using the BN speech recognizer. In the baseline system, this time-alignment was used for both the speech and the EMG signals. Because we have a marker channel in each signal, the marker signal is used to offset the two signals to get accurate time-synchronization. Then the AF training and testing procedures were applied both on the speech and the five-channel concatenated EMG signals, with the ST and E4 features. The averaged F-scores of all 29 AFs are 0.492 for EMG-ST, 0.686 for EMG-E4, and 0.814 for the speech signal. Fig. 10 shows individual AF performances for the speech and EMG signals along with the amount of training data in frames. The E4

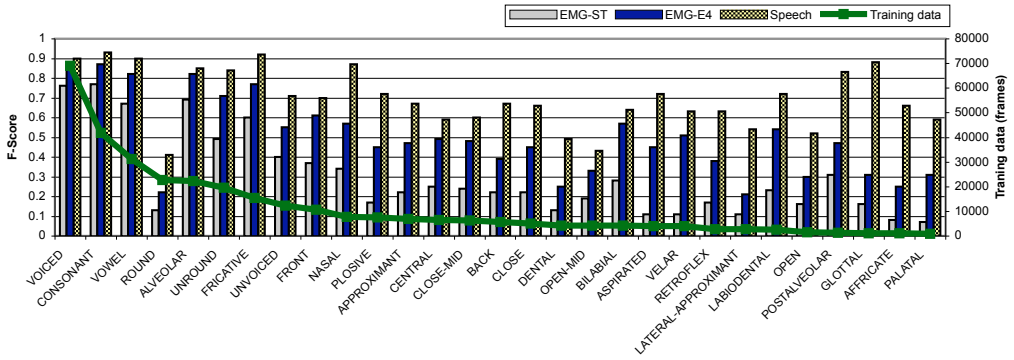


Figure 10: F-scores of the EMG-ST, EMG-E4 and speech articulatory features vs. the amount of training data.

significantly outperforms ST in that the EMG-E4 feature performance is much closer to the speech feature performance.

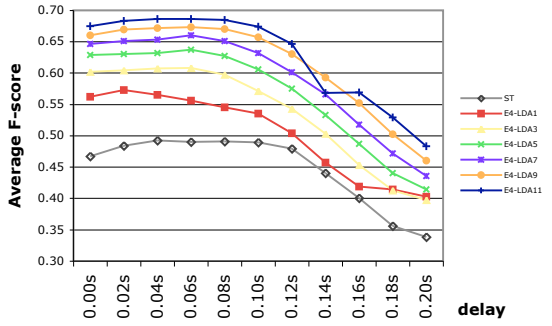


Figure 11: F-scores of concatenated five-channel EMG-ST and EMG-E4 articulatory features with various LDA frame sizes on time delays for modeling anticipatory effect.

We also conducted the time-delay experiments as done in previous ones to investigate the EMG vs. speech anticipatory effect. Fig. 11 shows the F-scores of E4 with various LDA frame sizes and delays. We observe similar anticipatory effect of E4-LDA and ST with time delay around 0.02 to 0.10 second. Compared to the 90-dimension ST feature, E4-LDA1 has a dimensionality of 25 while having a much higher F-score. The figure also shows that a wider LDA context width provides a higher F-score and is more robust for modeling the anticipatory effect, because LDA is able to pick up useful information from the wider context.

3.3.2 EMG Channel Pairs

In order to analyze E4 for individual EMG channels, we trained the AF classifiers on single channels and channel pairs. The F-scores are shown in Fig. 12. It shows E4 outperforms ST in all configurations. Moreover, E4 on single-channel EMG 1, 2, 3, 6 are already

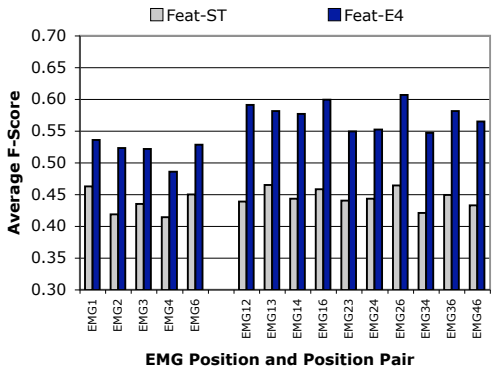


Figure 12: F-scores of the EMG-ST and EMG-E4 AFs on single EMG channel and paired EMG channels.

better than the all-channel ST's best F-score 0.492. For ST, the paired channel combination only provides marginal improvements; in contrast, for E4, the figure shows significant improvements of paired channels compared to single channels. We believe this significant improvements come from a better decorrelated feature space provided by E4.

3.3.3 Decoding in the Stream Architecture

We then conducted a full decoding experiment with the stream architecture. The test set was divided into two equally-sized subsets, on which the following procedure was done in two-fold cross-validation. On the development subset, we incrementally added the AF classifiers one by one into the decoder in a greedy approach, i.e., the AF that helps to achieve the best WER was kept in the streams for later experiments. After the WER improvement was saturated, we fixed the AF sequence and applied them on the test subset. Fig. 13 shows the WER and its relative improvements averaged on the two cross-validation turns. With five AFs, the WER tops 11.8% relative improvement, but

there is no additional gain with more AFs.

Among the selected AFs, only four of them are selected in both cross-validation turns. This inconsistency suggests a further investigation of AF selection is necessary for generalization.

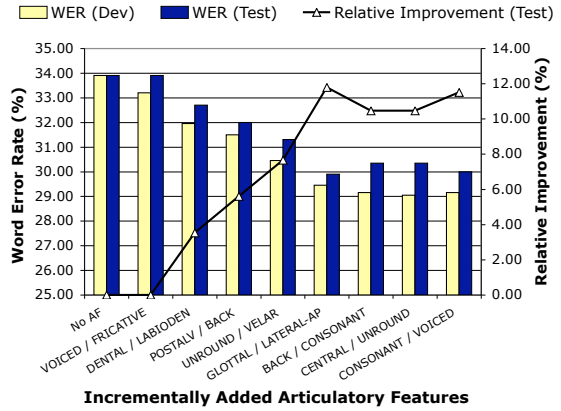


Figure 13: Word error rates and relative improvements of incrementally added EMG articulatory feature classifiers in the stream architecture. The two AF sequences correspond to the best AF-insertion on the development subsets in two-fold cross-validation.

4 COLLECTING MORE DATA

We are making efforts on larger-scale data collection of EMG speech. The targeted total number of speakers is in dozens and the recording modalities include acoustic speech, EMG, and video. Each speaker participates in two recording sessions, each of which includes a part of normal audible speech recording and a part of silent mouthing speech recording. In each part, two sets of phonetically balanced sentences are collected. One set is referred to as the general set and it exists in every part of every speaker. The other set is a speaker specific set, which is different for different speakers. Per part, the general set contains 10 sentences and the speaker specific set contains 40 sentences.

The data collection process is designed to be as unbiased as possible, e.g., to eliminate the fatigue factor. The two sessions are recorded one week apart. Besides, the order of the silent part and the audible part is reversed in the two sessions. In each recording part, the two sentence sets are mixed together into a set of 50 sentences and the sentences appear in random order. Table 4 shows the data details per speaker.

With this larger EMG corpus, we expect to be able to study the effects of speaker dependency, session dependency, audible versus mouthing speech kinematics, just to name a few.

Table 4: Data per speaker.

Speaker	
Session 1	Session 2
Part 1 audible speech rand(10+40 sentences)	Part 1 silent speech rand(10+40 sentences)
Part 2 silent speech rand(10+40 sentences)	Part 2 audible speech rand(10+40 sentences)

5 CONCLUSIONS

We have presented our recent advances on EMG speech recognition research, which has the advantages of better noise robustness and better applicability compared to traditional acoustic speech recognition. With the special EMG feature extraction methods and articulatory feature analyses, we have advanced the EMG speech recognition research from isolated word recognition to phone-based continuous speech recognition. Besides, the introduction of anticipatory effect modeling also plays an important role in this study. In summary, the EMG articulatory feature performance improves from 0.467 to 0.686 and the overall speech recognition word error rate improves from 86.8% to 29.9%.

This research topic is relatively new and unexplored with many questions waiting for answers. Although the proposed special EMG feature extraction methods do improve the performance, we believe they are still sub-optimal. Designing a better EMG feature extraction method for speech recognition is still an open problem and we are continuously working on it. Another issue is that the multi-channel EMG signals are inherently asynchronous with respect to articulatory apparatus movements. How to model this asynchronicity remains an open problem. We believe this modeling would benefit the study of speech recognition as well as articulatory kinematics.

REFERENCES

Becker, K. (2005). Varioport. *http://www.becker-meditec.de*.

Betts, B. and Jorgensen, C. (2006). Small vocabulary communication and control using surface electromyography in an acoustically noisy environment. In *Proc. HICSS*, Hawaii.

Chan, A., Englehart, K., Hudgins, B., and Lovely, D. (2002). Hidden Markov model classification of myoelectric signals in speech. *IEEE Engineering in Medicine and Biology Magazine*, 21(4):143–146.

Fromkin, V. and Ladefoged, P. (1966). Electromyography in speech research. *Phonetica*, 15.

Jorgensen, C. and Binsted, K. (2005). Web browser control using EMG based sub vocal speech recognition. In *Proc. HICSS*, Hawaii.

Jorgensen, C., Lee, D., and Agabon, S. (2003). Sub auditory speech recognition based on EMG signals. In *Proc. IJCNN*, Portland, Oregon.

Maier-Hein, L., Metze, F., Schultz, T., and Waibel, A. (2005). Session independent non-audible speech recognition using surface electromyography. In *Proc. ASRU*, San Juan, Puerto Rico.

Manabe, H., Hiraiwa, A., and Sugimura, T. (2003). Unvoiced speech recognition using EMG-Mime speech recognition. In *Proc. CHI*, Ft. Lauderdale, Florida.

Manabe, H. and Zhang, Z. (2004). Multi-stream HMM for EMG-based speech recognition. In *Proc. IEEE EMBS*, San Francisco, California.

Metze, F. and Waibel, A. (2002). A flexible stream architecture for ASR using articulatory features. In *Proc. ICSLP*, Denver, CO.

Yu, H. and Waibel, A. (2000). Streaming the front-end of a speech recognizer. In *Proc. ICSLP*, Beijing, China.

MRI SHOULDER COMPLEX SEGMENTATION AND CLASSIFICATION

Gabriela Pérez[1], J. F. Garamendi[2]
[1]*Departamento de Ciencias de la Computación,* [2]*Laboratorio de Imagen Médica y Biometría, Madrid, Spain*
gabriela.perez@urjc.es, juanfrancisco.garamendi@urjc.es

R. Montes Diez[3], E. Schiavi[4]
[3]*Departamento de Estadística e Investigación Operativa,* [4]*Departamento de Matemática Aplicada*
raquel.montes@urjc.es, emanuele.schiavi@urjc.es
Universidad Rey Juan Carlos, Madrid, Spain

Keywords: MRI, shoulder complex, segmentation, classification, multiphase Chan-Vese model.

Abstract: This paper deals with a segmentation (classification) problem which arises in the diagnostic and treatment of shoulder disorders. Classical techniques can be applied successfully to solve the binary problem but they do not provide a suitable method for the multiphase problem we consider. To this end we compare two different methods which have been applied successfully to other medical images modalities and structures. Our preliminary results suggest that a successful segmentation and classification has to be based on an hybrid method combining statistical and geometric information.

1 INTRODUCTION

Shoulder imaging is one of the major applications in MRI and the primary diagnostic tool in the evaluation of musculoskeletal disease, (Vahlensieck, 2000), (Ehman et al., 2001).

Accurate diagnosis and treatment of painful shoulder and others musculoskeletal complaints and disorders (such as arthritis, abnormalities, bone tumors, worn-out cartilage, torn ligaments, or infection) may prevent from functional loss, instability and disability. Recent interest is also in musculoskeletal tumor and disorders associated with HIV infection and AIDS, (Biviji et al., 2002), (Johnson and Steinbach LS, 2003). In order to provide a reliable method for successful clinical evaluation an increasing effort has to be done in mathematical engineering and biomedical imaging where the specific protocols of 2D segmentation, 3D reconstruction, feature extraction and 4D motion are modeled. In this approach for image guided analysis of shoulder pathologies, automatic and unsupervised segmentation and classification represent the first challenging task. In fact, practical difficulties arise due to the high resolution required for visualization of small but critical structures, to the gross inhomogeneities of field coil response, to the degree of noise present with the signal and to extreme low contrast details between some distinct anatomical structures (fat, bone regions, muscle and tendons, ligaments and cartilage). The existence of a general technique able to cope with all these difficulties for all 3D MRI images sequences is still an open question. A preliminary analysis of the model problem is done here, where a multiphase (2 phases, 4 classes) variational framework is considered for 2D image segmentation and classification. Notice that 2D segmentation is a fundamental step towards the 3D morpho-dynamic reconstruction problem of automatic segmentation. This in turn allows for motion tracking for 4D reconstruction and visualization of musculoskeletal structures.

2 MATERIAL AND METHODS

This contribution is devoted to the preliminary analysis and application of a modified multiphase segmentation and classification algorithm based on previous work of Chan and Vese (Chan and Vese, 2001). This multiphase approach can manage the classification problem underlying the segmentation exercise so broadening the scope of these PDE-based segmenta-

tion models.

In order to validate our results we compare with a mixture density estimation algorithm for image classification previously presented in (Mignotte et al., 2001), in the context of brain MRI images.
As an application of our method we consider coronal and transverse (axial), 2D MRI shoulder images extracted from two 3D sequences. The images are courtesy of the Ruber International Hospital in Madrid.

The shoulder joint is composed of three bones: the clavicle (collarbone), the scapula (shoulder blade), and the humerus (upper arm bone). The bones of the shoulder are held in place by muscles, tendons and ligaments. Tendons are tough cords of tissue that attach the shoulder muscles to bone and assist the muscles in moving the shoulder. Ligaments attach shoulder bones to each other, providing stability. The ends bones are covered by cartilage which provides painless motion. See Figure 1.

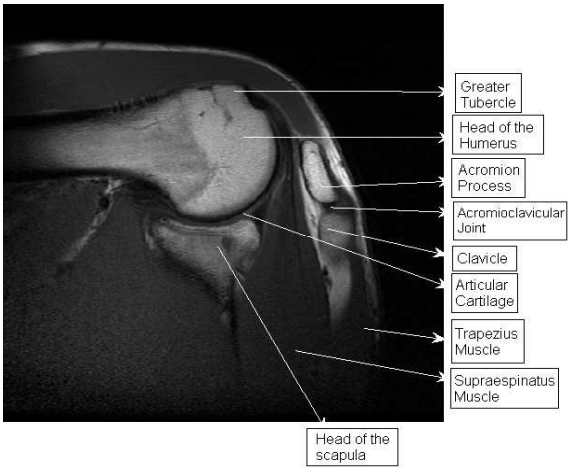


Figure 1: Components of the shoulder, Coronal MR image.

The classification problem we are about can be considered in the framework of minimal partition problems (Mumford-Shah) and cannot be dealt with classical techniques whereas binarization of image sequence is not suitable to produce the segmentation of all the classes we are interested in. Nervertheless it is interesting to compare the binary images obtained with thresholding techniques for the two classes (1 phase) problem in order to assess the performance of our algorithms when the full classification problem is considered. To cope with the difficulties above mentioned we consider two different approaches based on density mixture estimations (see (Mignotte et al., 2001)) and a variational model formulated in a level set framework (Chan and Vese, 2001).

The proposed algorithms are described in the next sections. The results obtained with classical (global and local thresholding or the popular k-means algorithm) techniques for the 2 classes (1 phase) problem are also reported for comparison. In particular we used the original Otsu's method (Otsu, 1979) and the Ridler-Calvard technique (Ridler and Calvard, 1978). We show the results obtained in figures 3 and 4 (d) and e)).

2.1 Density Mixture Estimation

In the case of the density mixture estimations framework the original magnitude images have been preprocessed in order to eliminate the high frequencies associated to noise and to increase the low contrast present in some parts of the image. As in (Brinkmann and Manduca, 1998), (Pérez et al., 2004). We consider a low pass homomorphic filter in the frequency domain which has been successfully used in previous works.

The initial pre-processing step is performed with a homomorphic filter in order to correct the gray scale inhomogeneity field. These inhomogeneities are known to appear in MR images as systematic changes in the local statistical characteristics of tissues and are often quite subtle to the human eye. However, even inhomogeneities that are invisible to the human observer alter tissue characteristics enough to hamper automated and semi-automated classification.

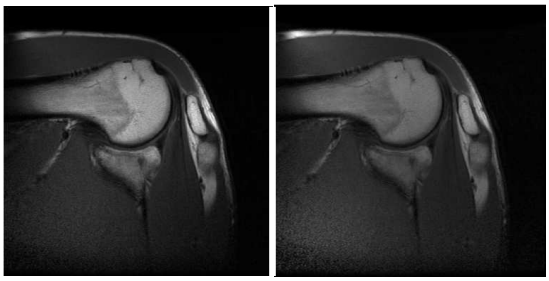


Figure 2: On the left the original image and on the right the pre-processed, corrected image.

Then, in a denoising step, the homogenized image is then filtered again with an adaptative filter to produce 2D wiener denoised sequence of the original image. The denoised slices are then normalized using a dynamical range operator in order to increase the (low) contrast present in the images. We then characterized the different soft tissues and bony structures in 4 classes (bone, muscle, cartilage, fat) partitioning the shoulder complex estimating their initial parameter statistics.

In order to show the basic steps of the algorithm we follow the Bayesian mixture parameter estimation method proposed by (Mignotte et al., 2001)

Let $Z = (X,Y)$ define two random fields where $Y = \{Y_s, s \in S\}$ represents the field of observations corresponding to the pixels of the MR image and $X = \{X_s, s \in S\}$ corresponds to the label field representing the segmented image. Following a Bayesian approach, the posterior distribution of (X,Y), $P(x|y)$, will result by combining the prior distribution, assumed to be stationary and Markovian,

$$P(x) = exp\{-\sum_{<s,t>} \beta(1-\delta(x_s,x_t))\},$$

and the site-wise likelihood $P(y_s|x_s)$, modelled as a mixture of densities

$$P(y_s|x_s,k,\Phi_k) = \sum_{k=1}^{K} \pi_k P(y_s|x_s,k,\Phi_k)$$

where π_k are the mixing proportion ($\sum_k \pi_k = 1$) and where $P(y_s|x_s,k,\Phi_k)$ define Gaussian distributions, with parameters $\Phi_k = (\mu_k, \sigma_k^2)$ in each segmented class k.

Let $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_K)$ and $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$. In order to proceed with the segmentation procedure, we perform the following algorithm:

0. Initialize parameters $(\Phi^{[0]}, \pi^{[0]})$.

Given $(\Phi^{[p]}, \pi^{[p]})$, we can calculate $(\Phi^{[p+1]}, \pi^{[p+1]})$ by

1. Using the Gibbs sampler, simulate one realization x from the posterior distribution $P(x|y)$ with parameter $(\Phi^{[p]}, \pi^{[p]})$.
2. Define $(\Phi^{[p+1]}, \pi^{[p+1]})$ as the ML estimation of the data (Y,x)
3. Repeat till convergence is achieved.

2.2 Active Contours Without Edges

Since the work of Kass (Kass et al., 1987) is well known that the segmentation problem of digital images can be dealt with in the framework of variational calculus. Nevertheless in medical images there are often no sharp-gradient induced edges at all and region-based active contours driven by gradients can fail in automatic approaches. Recently a new model has been suggested by Chan-Vese which can be deduced from the Mumford -Shah minimal partition problem, (Mumford and Shah, 1989), a basic problem in computer vision. Successful applications of this method have been reported in many papers and fields (see (Chan and Vese, 2001) and (Vese and Chan, 2002)). Our aim is to show that this active contour without edges model (or statistical feature driven model) can be used to solve the classification problem we consider here where a multiphase level set framework for image segmentation is implemented. The basic idea is that, fixed the number of classes in which we are interested in (fat, bone regions, muscle and tendons, ligaments and cartilage), it is sufficient to consider a two phase model, say ϕ_1, ϕ_2, in order to provided partition of the image in four classes (($\phi_1 > 0$ and $\phi_2 > 0$), ($\phi_1 < 0$ and $\phi_2 > 0$), ($\phi_1 > 0$ and $\phi_2 < 0$), ($\phi_1 < 0$ and $\phi_2 < 0$)).

Now, we explain the one phase (binary) and two phases models considered in the experiments. Let $\Omega \subset I\!R^2$ be an open, bounded domain (usually a square) where $(x,y) \in \Omega$ denotes pixel location and $I(x,y)$ is a function representing the intensity image values. Let moreover the level sets functions denoted by $\phi_1, \phi_2 : \Omega \to I\!R$. The union of the zero-level sets of ϕ_1 and ϕ_2 represents the edges of segmentation. Using this formalism the functions ϕ_1 and ϕ_2 can be characterized as the minimum of the following energy functional:

$$\begin{aligned} F(C,\Phi) &= \int_\Omega (I-c_{11})^2 H(\phi_1)H(\phi_2)dxdy \\ &+ \int_\Omega (I-c_{10})^2 H(\phi_1)(1-H(\phi_2))dxdy \\ &+ \int_\Omega (I-c_{01})^2 (1-H(\phi_1))H(\phi_2)dxdy \\ &+ \int_\Omega (I-c_{00})^2 (1-H(\phi_1))(1-H(\phi_2))dxdy \\ &+ \nu \int_\Omega |\nabla H(\phi_1)|dxdy + \\ &+ \nu \int_\Omega |\nabla H(\phi_2)|dxdy \end{aligned} \quad (1)$$

where $C = (c_{11}, c_{10}, c_{01}, c_{00})$ is a constant vector representing the mean of each region (or class), $\Phi = (\phi_1, \phi_2)$, ν is a parameter of smoothness and $H(x)$ is the Heaviside function, $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise,

The Euler-Lagrange equations obtained by minimizing (1) with respect to C and Φ are solved with a gradient descent method leading to the coupled parabolic PDE system (Vese and Chan, 2002):

$$\begin{aligned} \frac{\partial \phi_1}{\partial t} &= \delta_\varepsilon(\phi_1) \left\{ \nu \nabla \cdot \left(\frac{\nabla \phi_1}{|\nabla \phi_1|} \right) - \right. \\ &- \left[(I-c_{11})^2 - (I-c_{01}^2) \right] H(\phi_2) + \\ &+ \left. \left[(I-c_{10})^2 - (I-c_{00}^2) \right] (1-H(\phi_2)) \right\} \end{aligned} \quad (2)$$

$$\begin{aligned} \frac{\partial \phi_2}{\partial t} &= \delta_\varepsilon(\phi_2) \left\{ \nu \nabla \cdot \left(\frac{\nabla \phi_2}{|\nabla \phi_2|} \right) - \right. \\ &- \left[(I-c_{11})^2 - (I-c_{01}^2) \right] H(\phi_1) + \\ &+ \left. \left[(I-c_{10})^2 - (I-c_{00}^2) \right] (1-H(\phi_1)) \right\}. \end{aligned} \quad (3)$$

Where δ_ε denotes a smooth (not compactly supported) approximation to the Dirac delta distribution. Notice that the equations (2) and (3) are (weakly) coupled in the lower order terms. In case of two regions only one level set function ϕ is needed. The resulting one phase energy functional to minimize is as follows:

$$\begin{aligned} E_{cv}(c_1,c_0,\phi) &= \nu\int_\Omega |\nabla H(\phi)|dxdy + \\ &+ \int_\Omega H(\phi)|I(x,y)-c_1|^2dxdy + \\ &+ \int_\Omega (1-H(\phi))|I(x,y)-c_0|^2dxdy \end{aligned} \tag{4}$$

and the associated gradient descent equation is :

$$\begin{aligned} \frac{\partial\phi}{\partial t} &= \delta_\varepsilon(\phi)\left[\nu\nabla\cdot\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \right. \\ &- \left. |I(x,y)-c_1|^2+|I(x,y)-c_0|^2\right]. \end{aligned} \tag{5}$$

The equations (2), (3) or (5) have to be complemented with feasible (due to the non-uniqueness of the corresponding steady states) initial conditions and homogeneous boundary conditions of Neumann type (no flux). As in Chan and Vese (Chan and Vese, 2001) the steady states associated to system (2), (3) or the eq. (5) can be asymptotically reached by using a gradient descent method where δ_ε is substituted by 1 (this is possible because δ_ε has no compact support). Numerically, as we are concerned with the quality of the classification and not in to speed it up, we used a simple first order (in time) Euler explicit finite difference scheme and weighted, centered, second order formulas in space, with a regularization of the (degenerate) diffusion term to avoid division by zero (which occurs in homogeneous, very low gradient regions which are located far from the active contour and do not affect the final segmentation as soon as the regularizing parameter is small). The time steps have been choosen accordingly in order to preserv numerical stability and convergence.

3 RESULTS

We present the results obtained by applying the above methods to a pair of slices extracted from a volume MRI sequence of the shoulder complex. The slices dimensions are 512x512.

Binary segmentations obtained with both methods (the bayesian density mixture estimation and the PDE-based hybrid active contours method without edges) are shown in Figures 3-4 before of the multiphase classification, see figures 5-6.

Figure 3: Slice 1. Segmentation image for one phase (2 classes) with: a) k-means b) Density mixture c) Active contours without edges d) Otsu's and e) Ridler Calvard algorithms.

For comparison and in both cases, we also include the results provided by classical methods. Binary segmentation is also used to assess the parameters involved in the model equations and to provide automatic, robust initial conditions for the evolutive problem in the multiphase case.

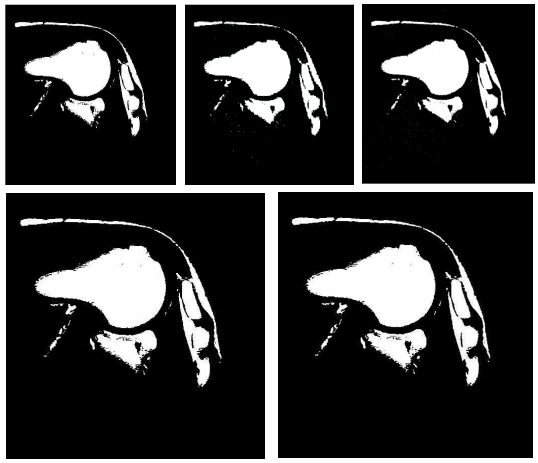


Figure 4: Slice 2. Segmentation image for one phase (2 classes) with: a) k-means b) Density mixture c) Active contours without edges d) Otsu's and e) Ridler Calvard algorithms.

In Figures 3-4 we see that in both cases the bony structures (head of scapula, head of humerus, clavicle, acromion) are properly classified. Background, skin, and muscle are also characterized in the binary images as the soft (tissue) class. Visual inspection

suggests convergence to the same limit solution. This is indeed confirmed when the differences images are computed and classical methods (first row) are compared.

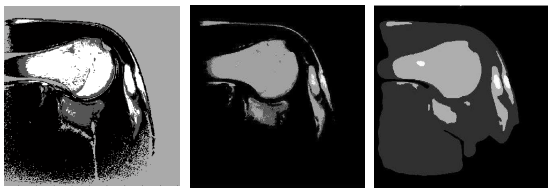


Figure 5: Slice 1. Segmentation image for two phases (4 classes) a) k-means b) Density mixture c) Active contours without edges algorithms.

As aspected, some more differences between the quality reconstruction of the different methods can be appreciated in the multiphase (four classes) classification problem. In Figures 5-6 we report the results obtained with the classical (k-means) algorithm (left), the bayesian mixture model (center) and the Chan-Vese model (right). The greater tubercle and the head of the humerus are properly classified and shaped with our methods (center and right) while the classical k-means fails in both aspects (and in both slices, see Figures 5-6, on the left, where the bone is under-estimated and muscle is wrongly detected). Articular cartilage has been detected in (center and right) but not in (left). Muscle is properly classified with the Chan-Vese model (right) and the classical method (left) but no classification has been done in the bayesian approach where the background is assigned to the same class. At the same time the head of the scapula has been properly classified in (right) but not in (center) where the shape, nevertheless is correctly obtained. Notice also that the acromial process has been characterized by the two methods.

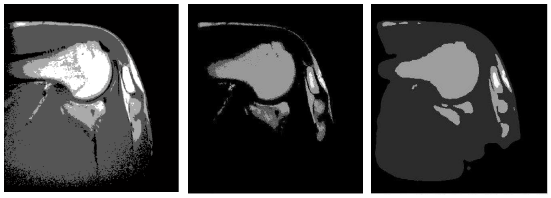


Figure 6: Slice 2. Segmentation image for two phases (4 classes) a) k-means b) Density mixture c) Active contours without edges algorithms.

4 CONCLUSIONS

We considered the problem of automatic segmentation of 2D images using an hybrid, statistical and geometrical model based on Chan-Vese work. This method provides correct classification of bony structures but soft tissues are not yet properly classified. This is also manifested in the bayesian approach. The differences between the results obtained with the two methods suggest the conclusion that hybrid methods can give better results as far as the right statistics are included in the model and this will be the aim of our future work.

ACKNOWLEDGEMENTS

This work was partially granted by ”Research on Molecular and Multimodality Medical Imaging Network” of the Carlos III Health Institute. Finally the authors of the present work would like to thank Ruber International Hospital and the neuroradiologist Dr. Juan Linera.

REFERENCES

Biviji, A., Paiement, G., Davidsion, A., and Steinbach, L. (2002). Musculoeskeletal manifestations of the human immunodeficiency virus infection. *Of the American Academy of Orthopedic Surgeons*, pages 10:312–20.

Brinkmann, B. and Manduca, A. (1998). Optimized homomorphic unsharp masking for mr grayscale inhomogeneity correction. In *IEEE transactions on medical imaging, p. 62-66*, volume 17.

Chan, T. F. and Vese, L. A. (2001). Active contours without edges. ieee transactions on image processing. In *IEEE Transactions on Image Processing*, volume 10.

Ehman, R., Megibow, A., MacCauley, T., Bluemke, D., and Steinbach, L. (2001). Musculoskeletal imaging. In *the 24th Annual Course of the Society of Computed Body Tomography and Magnetic Resonance*, Miami.

Johnson, R. and Steinbach LS, e. (2003). Essentials of musculoskeletal imaging. american academy of orthopedic surgeons. Chicago.

Kass, M., WitKin, A., and Terzopoulos, D. (1987.). Snakes: Active contour models. In *Intl. J. Comput. Vision, 1:321-331*.

Mignotte, M., Meunier, J., Soucy, J. P., and Janicki, C. (2001). classification of brain spect images using 3d markov random field and density mixture estimations. 5th world multi-conference on systemics, cybernetics and informatics. In *Concepts and Applications of Systemics and Informatics*, volume 10, pages 239–244, Orlando.

Mumford, D. and Shah, J. (1989). Optimal approximation by piecewise smooth functions and associated variational problems. In *Communications on Pure Applied Mathematics, p. 577-685*, volume 42.

Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Pérez, G., Diez, R. M., Hernández, J. A., and José San Martín (2004). A new approach to automatic segmentation of bone in medical magnetic resonance imaging. In *5th International Symposium, ISBMDA*, pages 21–26, Spain.

Ridler, T. and Calvard, S. (1978). Picture thresholding using an iterative selection method. *IEEE transactions on systems, man, and cybernetics*, 8(8).

Vahlensieck, M. (2000). Mri of the shoulder. *European Radiology*, 10(2):242–249.

Vese, L. A. and Chan, T. F. (2002). A multiphase level set framework for image segmentation using the mumford and shah model. In *International Journal of Computer Vision, p. 271-293*, volume 50.

EVOLUTIONARY COMPUTATION APPROACH TO ECG SIGNAL CLASSIFICATION

Farid Melgani
Dept. of Information and Communicatrion technologies, University of Trento, Via Sommarive, 14-I-38050, Trento, Italy
melgani@dit.unitn.it

Yakoub Bazi
College of Engineering, Al Jouf University, 2014 Sakaka, Saudi Arabia
yakoub.bazi@ju.edu.sa

Keywords: ECG classification, feature reduction, particle swarm optimization, support vector machine.

Abstract: In this paper, we propose a novel classification system for ECG signals based on particle swarm optimization (PSO). The main objective of this system is to optimize the performance of the support vector machine (SVM) classifier in terms of accuracy by automatically: i) searching for the best subset of features where to carry out the classification task; and ii) solving the SVM model selection issue. Experiments conducted on the basis of ECG data from the MIT-BIH arrhythmia database to classify five kinds of abnormal waveforms and normal beats confirm the effectiveness of the proposed PSO-SVM classification system.

1 INTRODUCTION

The recent literature reports different and interesting methodologies for the automatic classification of electrocardiogram (ECG) signals (e.g., de Chazal et Reilly, 2006, and Inan et Giovangrandi, 2006). However, in the design of an ECG classification system, there are still some open issues, which if suitably addressed may lead to the development of more robust and efficient classifiers. One of these issues is related to the choice of the classification approach to be adopted. In particular, we think that, despite its great potential, the SVM approach has not received the attention it deserves in the ECG classification literature compared to other research fields. Indeed, the SVM classifier exhibits a promising generalization capability thanks to the maximal margin principle (MMP) it is based on (Vapnik, 1998). Another important property is its lower sensitivity to the curse of dimensionality compared to traditional classification approaches. This is explained by the fact that the MMP makes unnecessary to estimate explicitly the statistical distributions of classes in the hyperdimensional feature space in order to carry out the classification task. Thanks to these interesting properties, the SVM classifier proved successful in numerous and different application fields, such as 3D object recognition (Pontil et Verri, 1998), biomedical imaging (El-Naqa et al., 2002), remote sensing (Melgani et Bruzzone, 2004 and Bazi et Melgani, 2006). Turning back to ECG classification, other issues which can be identified are: 1) feature selection is not performed in a completely automatic way; and 2) the selection of the best free parameters of the adopted classifier is generally performed empirically (model selection issue).

In this paper, we present in a first step a thorough experimental exploration of the SVM capabilities for ECG classification. In a second step, in order to address the aforementioned issues, we propose to optimize further the performances of the SVM approach in terms of classification accuracy by 1) automatically detecting the best discriminating features from the whole considered feature space and 2) solving the model selection issue. The detection process is implemented through a particle swarm optimization (PSO) framework that exploits a criterion intrinsically related to the SVM classifier properties, namely the number of support vectors (#SV).

2 PROPOSED APPROACH

2.1 Support Vector Machine (SVM)

Let us first for simplicity consider a supervised binary classification problem. Let us assume that the training set consists of N vectors $\mathbf{x_i} \in \Re^d$ (i = 1, 2, …, N) from the d-dimensional feature space **X**. To each vector $\mathbf{x_i}$, we associate a target $y_i \in \{-1, +1\}$. The linear SVM classification approach consists of looking for a separation between the two classes in **X** by means of an optimal hyperplane that maximizes the separating margin (Vapnik, 1998). In the nonlinear case, which is the most commonly used as data are often linearly nonseparable, they are first mapped with a kernel method in a higher dimensional feature space, i.e., $\Phi(X) \in \Re^{d'}$ (d'> d). The membership decision rule is based on the function sign[f(**x**)], where f(**x**) represents the discriminant function associated with the hyperplane in the transformed space and is defined as:

$$f(\mathbf{x}) = \mathbf{w}^* \cdot \Phi(\mathbf{x}) + b^* \quad (1)$$

The optimal hyperplane defined by the weight vector $\boldsymbol{w}^* \in \Re^{d'}$ and the bias $b^* \in \Re$ is the one that minimizes a cost function that expresses a combination of two criteria: margin maximization and error minimization. It is expressed as:

$$\Psi(\boldsymbol{w},\boldsymbol{\xi}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i \quad (2)$$

This cost function minimization is subject to the following constraints:

$$y_i(\boldsymbol{w}\cdot\Phi(\boldsymbol{x}_i)+b) \geq 1-\xi_i, \; i = 1,\ldots, N \quad (3)$$

and

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, N \quad (4)$$

where the ξ_i's are slack variables introduced to account for non-separable data. The constant C represents a regularization parameter that allows to control the shape of the discriminant function. The above optimization problem can be reformulated through a Lagrange functional, for which the Lagrange multipliers can be found by means of a dual optimization leading to a Quadratic Programming (QP) solution, i.e.,

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j) \quad (5)$$

under the constraints:

$$\alpha_i \geq 0 \text{ for } i = 1, 2, \ldots, N \quad (6)$$

and

$$\sum_{i=1}^{N}\alpha_i y_i = 0 \quad (7)$$

where $\boldsymbol{\alpha}=[\alpha_1, \alpha_2, \ldots, \alpha_N]$ is the vector of Lagrange multipliers and $K(\cdot,\cdot)$ is a kernel function. The final result is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space ***X***:

$$f(\boldsymbol{x}) = \sum_{i\in S}\alpha_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x}) + b^* \quad (8)$$

The set S is a subset of the indices {1, 2, …, N} corresponding to the non-zero Lagrange multipliers α_i's, which define the so-called support vectors.

As described above, SVMs are intrinsically binary classifiers. But the classification of ECG signals often involves the simultaneous discrimination of numerous information classes. In order to face this issue, different multiclass classification strategies can be adopted (Melgani et Bruzzone, 2004). In this paper, we shall consider the commonly used one-against-all strategy.

2.2 PSO Principles

Particle swarm optimization (PSO) is a stochastic optimization technique introduced recently by Kennedy and Eberhart, which is inspired by social behavior of bird flocking and fish schooling (Kennedy et Eberhart, 2001). Similarly to other evolutionary computation algorithms such as genetic algorithms (GAs) (Bazi et Melgani, 2006), PSO is a population-based search method, which exploits the concept of social sharing of information. This means that each individual (called *particle*) of a given population (called *swarm*) can profit from the previous experiences of all other individuals from the same population. During the iterative search process in the d-dimensional solution space, each particle (i.e., candidate solution) will adjust its flying velocity and position according to its own flying experience as well as the experiences of the other companion particles of the swarm.

Let us consider a swarm of size S. Each particle P_i $(i = 1, 2, \ldots, S)$ from the swarm is

characterized by: 1) its current position $\mathbf{p}_i(t) \in \Re^d$, which refers to a candidate solution of the optimization problem at iteration t; 2) its velocity $\mathbf{v}_i(t) \in \Re^d$; and 3) the best position $\mathbf{p}_{bi}(t) \in \Re^d$ identified during its past trajectory. Let $\mathbf{p}_g(t) \in \Re^d$ be the best global position found over all trajectories traveled by the particles of the swarm. The position optimality is measured by means of one or more fitness functions defined in relation to the considered optimization problem. During the search process, the particles move according to the following equations:

$$\begin{cases} \mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + c_1 \cdot r_1(t)\left(\mathbf{p}_{bi}(t) - \mathbf{p}_i(t)\right) \\ \qquad + c_2 \cdot r_2(t)\left(\mathbf{p}_g(t) - \mathbf{p}_i(t)\right) \\ \mathbf{p}_i(t+1) = \mathbf{p}_i(t) + \mathbf{v}_i(t) \end{cases} \quad (9)\ (10)$$

where $r_1(\cdot)$ and $r_2(\cdot)$ are random variables drawn from a uniform distribution in the range [0, 1] so that to provide a stochastic weighting of the different components participating in the particle velocity definition. c_1 and c_2 are two acceleration constants regulating the relative velocities with respect to the best global and local positions, respectively. The inertia weight w is used as a tradeoff between global and local exploration capabilities of the swarm.

2.3 PSO Setup

The position $\mathbf{p}_i \in \Re^{d+2}$ of each particle P_i from the swarm is viewed as a vector encoding: 1) a candidate subset F of features among the d available input features, and 2) the value of the two SVM classifier parameters, which are the regularization and the kernel parameters C and γ, respectively. Since the first part of the position vector implements a feature detection task, each component (coordinate) of this part will assume either a "0" (feature discarded) or a "1" (feature selected) value. The conversion from real to binary values will be done by a simple thresholding operation at the 0.5 value.

Let $f(i)$ be the fitness function value associated with the ith particle P_i. The choice of the fitness function is important since, on its basis, the PSO evaluates the goodness of each candidate solution $\mathbf{p}_i$ for designing our SVM classification system. A possible choice is to adopt the class of criteria that estimates the leave-one-out error bound, which exhibits the interesting property of representing an unbiased estimation of the generalization performance of classifiers. In particular, for SVM classifiers, different measures of this error bound have been derived, such as the radius-margin bound and the simple support vector (SV) count (Vapnik, 1998). In this paper, we will explore the simple SV count as fitness criterion in the PSO optimization framework because of its simplicity and effectiveness as shown in the context of the classification of hyperspectral remote sensing images (Bazi et Melgani, 2006).

2.4 SVM Classification with PSO

- ***Initialization***

***Step 1*:** Generate randomly an initial swarm of size S.
Step 2: Set to zero the velocity vectors $\mathbf{v}_i$ (i= 1, 2,..., S) associated with the S particles.
***Step 3*:** For each position $\mathbf{p}_i \in \Re^{d+2}$ of the particle P_i $(i = 1, 2, ..., S)$ from the swarm, train an SVM classifier and compute the corresponding fitness function *f(i)* (i.e., the #SV measure).
Step 4: Set the best position of each particle with its initial position, i.e.,

$$\mathbf{p}_{bi} = \mathbf{p}_i, (i=1, 2,.., S) \quad (11)$$

- ***Search process***

Step 5: Detect the best global position $\mathbf{p}_\mathbf{g}$ in the swarm exhibiting the minimal value of the considered fitness function over all explored trajectories.
Step 6: Update the speed of each particle using Equation (9).
Step 7: Update the position of each particle using Equation (10). If a particle goes beyond the predefined boundaries of the search space, truncate the updating by setting the position of the particle at the space boundary and reverse its search direction (i.e., multiply its speed vector by -1). This will permit to forbid the particles from further attempting to go outside the allowed search space.
Step 8: For each candidate particle $\mathbf{p}_i$ (i= 1, 2,..., S), train an SVM classifier and compute the corresponding fitness function.
Step 9: Update the best position $\mathbf{p}_{bi}$ of each particle if its new current position $\mathbf{p}_i$ (i= 1, 2,..., S) has a smaller fitness function.

- ***Convergence***

Step 10: If the maximal number of iterations is not yet reached, return to *Step 5*.

- ***Classification***

Step 11: Select the best global position $\mathbf{p}_\mathbf{g}^*$ in the swarm and train an SVM classifier fed with the subset of detected features mapped by $\mathbf{p}_\mathbf{g}^*$ and modeled with the values of the two parameters C and γ encoded in the same position.
Step 12: Classify the ECG signals with the trained SVM classifier.

3 EXPERIMENTAL RESULTS

Our experiments were conducted on the basis of ECG data from the MIT-BIH arrhythmia database (Mark et Moody, 1997). In particular, the considered beats make reference to the following classes: normal sinus rhythm (N), atrial premature beat (A), ventricular premature beat (V), right bundle branch block (RB), left bundle branch block (LB), and paced beat (/). Similarly to (Inan et al., 2006), the beats were selected from the recordings of 18 patients, which correspond to the following files: 100, 102, 104, 105, 106, 107, 118, 119, 200, 201, 203, 205, 208, 212, 213, 214, 215, and 217. For feeding the classification process, we adopted in this study the two following kinds of features: i) ECG morphology features; and ii) three ECG temporal features that are the QRS complex duration, the RR interval (i.e., time span between two consecutive R points representing the distance between the QRS peaks of the present and previous beats), and the RR interval averaged over the ten last beats (de Chazal et Reilly, 2006). The total number of morphology and temporal features is equal to 303 for each beat. In order to train the classification process and to assess its accuracy, we selected randomly from the considered recordings 500 beats for the training set, whereas 42185 beats were used as test set (thus, the training set represents just 1.18% of the test set). The detailed numbers of training and test beats are reported for each class in Table 1. Classification performance was evaluated in terms of three accuracy measures, which are: 1) the overall accuracy (OA); 2) the accuracy of each class; and 3) the average accuracy (AA).

Due to the good performances generally achieved by the nonlinear SVM classifier based on the Gaussian kernel [6], we adopted this kernel in all experiments. The parameters C and γ were varied in the ranges $[10^{-3}, 200]$ and $[10^{-3}, 2]$, respectively. The k value and the number of hidden nodes (h) of the kNN and the RBF classifiers were tuned in the intervals [1, 15] and [10, 60], respectively. Concerning the PSO algorithm, we considered the following standard parameters: swarm size S=40, inertia weight w=0.4, acceleration constants c_1 and c_2 equal to the unity, and maximum number of iterations fixed to 40.

3.1 Experiment 1: Classification in the Original Feature Space

In this experiment, we applied the SVM classifier directly on the whole original hyperdimensional feature space which is composed of 303 features. During the training phase, the SVM parameters (i.e., C and γ) were selected according to a m-fold cross-validation (CV) procedure. In all experiments reported in this paper, we adopted a 5-fold CV. The same procedure was adopted to find the best parameters for the kNN and the RBF classifiers. The best values obtained for the three investigated classifiers are C=25, γ=0.5, k=3 and h=20. As reported in Table 2, the OA and AA accuracies achieved by the SVM classifier on the test set are equal to 87.95% and 87.60%, respectively. These results are much better than those achieved by the RBF and the kNN classifiers. Indeed, the OA and AA accuracies are equal to 82.78% and 82.34% for the RBF classifier, and 78.21% and 79.34% for the kNN classifier, respectively. This experiment appears to confirm what observed in other application fields, i.e., the superiority of SVM with respect to traditional classifiers when dealing with feature spaces of very high dimensionality.

3.2 Experiment 2: Classification based on Feature Reduction

In this experiment, we trained the SVM classifier in feature subspaces of various dimensionalities. The desired number of features was varied from 10 to 50 with a step of 10, namely from small to high dimensional feature subspaces. Feature reduction was achieved by means of the traditional Principal Component Analysis (PCA) algorithm. Figure 1-a depicts the results obtained in terms of OA by the three considered classifiers combined with the PCA algorithm, namely the PCA-SVM, the PCA-RBF and the PCA-kNN classifiers. In particular, it can be seen that, for all feature subspace dimensionalities except the lowest one (i.e., 10 features), the PCA-SVM classifier maintains a clear superiority over the two other classifiers. Its best accuracy was found by using a feature subspace composed of the first 40 components. The corresponding OA and AA accuracies are equal to 88.98% and 88%, respectively. Comparing these results with those obtained by the SVM classifier in the original feature space (i.e., without feature reduction), a slight increase of 1.03% and 0.4% in terms of OA and AA, respectively, was achieved (see Table 2).

3.3 Experiment 3: Classification with PSO-SVM

In this experiment, we applied the PSO-SVM classifier on the available training beats. At

convergence of the optimization process, we assessed the PSO-SVM classifier accuracy on the test samples. The achieved overall and average accuracies are equal to 91.44% and 91.19% corresponding to substantial accuracy gains with respect to what yielded either by the SVM classifier applied on all available features (+3.49% and +3.59%, respectively) or by the PCA-SVM classifier (+2.46% and +3.19%, respectively) (see Table 2 and Figure 1). Its worst class accuracy was obtained for atrial premature beats (A) (88.16%) while that of the SVM and the PCA-SVM classifiers corresponded to paced beats (/) (73.43%) and ventricular premature beats (V) (78.06%), respectively. This shows the capability of the PSO-SVM classifier to reduce significantly the gap between the worst and best class accuracies (8.25% against 15.43% and 20.21% for the PCA-SVM and the SVM classifiers, respectively) while keeping the overall accuracy to a high level.

4 CONCLUSIONS

The main novelty of this paper is to be found in the proposed PSO-based approach that aims at optimizing the performances of SVM classifiers in terms of classification accuracy by detecting the best subset of available features and by solving the tricky model selection issue. Its completely automatic nature renders it particularly useful and attractive. The results confirm that the PSO-SVM classification system boosts up significantly the generalization capability achievable with the SVM classifier. Finally, it is noteworthy that the general nature of the proposed PSO-SVM system makes it applicable not just to morphology and temporal features but to other kinds of features such as those based on wavelets and high-order statistics. Finally, other optimization criteria could be considered as well individually or jointly depending on the application requirements.

REFERENCES

Bazi Y., Melgani F., (2006). Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens., 44*, 3374-3385.

De Chazal F., Reilly R.B., (2006). A patient adapting heart beat classifier using ECG morphology and heartbeat interval features. *IEEE Trans. Biomedical Engineering, 43*, 2535-2543.

El-Naqa I., Yongyi Y., Wernick M.N., Galatsanos N.P., Nishikawa R.M., (2002). A support vector machine approach for detection of microcalcifications. *IEEE Trans. on Medical Imaging*, *21*, 1552-1563.

Inan O.T., Giovangrandi L., Kovacs J.T.A., (2006). Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Trans. Biomedical Engineering, 53*, 2507-2515.

Kennedy J., Eberhart R.C., (2001). *Swarm Intelligence*. San Mateo, CA: Morgan Kaufmann.

Mark R., Moody G., (1997). MIT-BIH Arrhythmia Database 1997 [Online]. Available: http://ecg.mit.edu/dbinfo.html.

Melgani F., Bruzzone L., (2004). Classification of hyperspectral remote sensing images with support vector machine. *IEEE Trans. on Geosci. Remote Sens.,* 42, 1778-1790.

Pontil M., Verri A., (1998). Support vector machines for 3D object recognition**",** *IEEE Trans. on Pattern Analysis and Machine Intelligence, 20*, 637-646.

Vapnik V., (1998). *Statistical Learning Theory*. New York: Wiley.

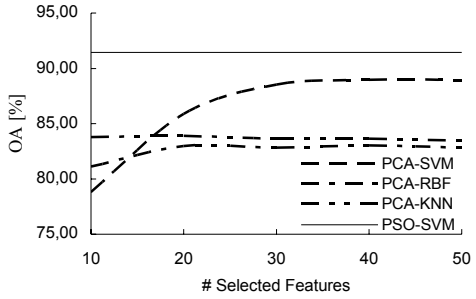


Figure 1: Overall percentage accuracy (OA) versus number of selected features (first principal components) achieved on the test beats by the different classifiers.

Table 1: Numbers of training and test beats used in the experiments.

Class	N	A	V	RB	/	LB	Total
Training beats	150	100	100	50	50	50	500
Test beats	24966	119	4239	3939	6971	1951	42185

Table 2: Overall (OA), average (AA), and class percentage accuracies achieved on the test beats by the different investigated classifiers.

Method	OA	AA	N	A	V	RB	/	LB
SVM	87.95	87.60	90.05	83.19	92.12	93.15	73.43	93.64
RBF	82.78	82.34	85.14	78.99	90.39	86.74	66.53	86.26
kNN	78.21	79.34	76.50	66.38	71.99	93.27	75.92	92.00
PCA-SVM	88.98	88.00	89.36	83.19	78.06	93.50	90.60	93.28
PCA-RBF	83.04	82.11	85.86	80.67	87.85	83.87	68.85	85.54
PCA-kNN	83.91	82.02	85.62	69.74	79.05	93.04	73.89	90.77
PSO-SVM	**91.44**	**91.19**	**91.12**	**88.16**	**93.70**	**93.70**	**92.01**	**96.41**

FEASABILITY OF YEAST AND BACTERIA IDENTIFICATION USING UV-VIS-SWNIR DIFUSIVE REFLECTANCE SPECTROSCOPY

J. S. Silva, R. C. Martins, A. A. Vicente and J. A. Teixeira
IBB - Institute for Biotechnology and BioEngineering, Universidade do Minho
Campus de Gualtar, 4710-057 Braga, Portugal
rui.martins@deb.uminho.pt

Keywords: Yeast, bacteria, UV-VIS-SWNIR reflectance spectroscopy, Singular value decomposition, Classification.

Abstract: UV-VIS spectroscopy is a powerfull qualitative and quantitative technique used in analytical chemistry, which gives information about electronic transitions of electrons in molecular orbitals. As in UV-VIS spectra there is no direct information on characteristic organic groups, vibrational spectroscopy (e.g. infrared) has been preferred for biological applications. In this research, we try to use state-of-the-art fiber optics probes to obtain UV-VIS-SWNIR diffusive reflectance measurements of yeasts and bacteria colonies on plate count agar in the region of 200-1200nm; in order to discriminate the following microorganisms: i) yeasts: *Saccharomyces cerevisiae*, *Saccharomyces bayanus*, *Candida albicans*, *Yarrowia lipolytica*; and ii) bacteria: *Micrococcus luteus*, *Pseudomonas fluorescens*, *Escherichia coli*, *Bacillus cereus*. Spectroscopy results show that UV-VIS-SWNIR has great potential for identifying microorganisms on plate count agar. Scattering artifacts of both colonies and plate count agar can be significantly removed using a robust mean scattering algorithm, allowing also better discriminations between the scores obtained by singular value decomposition. Hierarchical clustering analysis of UV-VIS and VIS-SWNIR decomposed spectral scores lead to the conclusion that the use of VIS-SWNIR light source produces higher discrimination ratios for all the studied microorganisms, presenting great potential for developing biotechnology applications.

1 INTRODUCTION

Spectroscopy is a powerful tool for biological applications, being applicable to liquids, solutions, pastes, powders, films, fibres, gases and surfaces, and making possible to characterize proteins, peptides, lipids, membranes, carbohydrates in pharmaceuticals, foods, plants or animal tissues (Hammes, 2005).

One of the most popular method is Infrared Spectroscopy (IR), and was firstly applied to biological materials in 1911 (Riddle et al., 1956). In the 1950s and 1960s research IR spectroscopy began to be applied for microorganism differentiation, but this research was abandoned due to the unsatisfactory results obtained with dispersive spectrometers (Dziuba et al., 2007). These were ignored during 20 years, until modern interferometric Fourier Transform Infra-Red spectrometers (FT-IR) and statistical computing methodologies became available (Dziuba et al., 2007).

Recent techniques using FT-IR allowed microbiological characterization and the discrimination at level of sorting better species and strains. Attenuated total reflection and IR micro-spectroscopy have been associated to the discrimination and identification of strains according to taxonomic classification, gram +/- factor, or even susceptibility to antibiotics and grown medium (Mariey et al., 2001).

FT-IR has also been used to identify lactic acid bacteria strains (e.g. *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus* and *Streptococcus* (Dziuba et al., 2007), and the rapid identification of *Acinetobacter* species (Winder et al., 2004). An extensive FT-IR spectroscopy database for the identification of bacteria from the two suborders *Micrococcineae* and *Corynebacterineae* (*Actinomycetales*, *Actinobacteria*) as well as other morphologically similar genera was established in 2002 by Helene Oberreuter and its team (Oberreuter et al., 2002). Furthemore, FT-IR was used for the first time to determine the ratios of different yeast species (*Saccharomyces cerevisiae*, *Hanseniaspora uvarum*) and two yoghurt lactic acid bacteria (*Lactobacillus acidophilus*, *Streptococcus salivarius ssp. thermophilus*) in suspensions

Table 1: Studied microorganisms characteristics and experimental conditions: Gram factor, colony colour and shape and integration time.

					Integration Time (ms)	
Microorganism	Gram	Colony color	Shape	Medium	UV-VIS (ms)	VIS-NIR (ms)
Saccharomyces cerevisiae	na	white	spherical	YPD	70	66
Saccharomyces bayanus	na	white	spherical	YPD	45	61
Candida albicans	na	white	spherical	YPD	70	66
Yarrowia lipolytica	na	white	rod	YPD	70	61
Micrococcus luteus	-	yellow	spherical	TSA	20	19
Pseudomonas fluorescens	-	translucent	rod	MP	20	30
Escherichia coli	-	translucent	rod	LB	30	13
Bacillus cereus	+	opaque	rod	LB	36	61

of distilled water (Oberreuter et al., 2000).

Raman Spectroscopy has also shown great potential for microorganisms identifications in microscopy, such as for *Candida* yeast strains and bacterial strains such as *Staphylococus*, *Enterococus* and *Echerichia Coli*) (Guibeta et al.,). Applications are also found in oral hygiene, for the identification of *Streptococcus* mutants, *S. sanguis*, *S.intermedius* and *S. oralis* (Berger and Zhu, 2003). Moreover, this technique is currently used to identify baterials cells of *Staphylococcus* under different cultivation conditions (Harz et al., 2005) and single yeast cells (Rch et al., 2005). Fluorescence Spectroscopy (FS) is one of the most important spectroscopic techniques in molecular biology, and consequently can also be used to microbiological identification. FS applications can be found on the differentiation of yeast and bacteria, by their intrinsic fluorescence to UV excitation (Bhatta et al., 2005).

UV-VIS-SWNIR spectroscopy is one of the most widely used techniques in analytical chemistry, but it has almost not been used for microorganism identification. This is perhaps attributed to the fact that UV-VIS spectroscopy records transmitions between electron energy levels from molecular orbitals, instead of vibrational or structural oscillation of molecular groups as in the infrared region. It is widely accepted that vibrational spectroscopy is more adequate for organic chemistry measurement than transitional spectroscopy. Nevertheless an asset of this technique has never been done use for microbiological identification.

Electronic transitions in the UV-VIS region depend upon the energy involved. For any molecular bound (sharing a pair of electrons), orbitals are a mixture of two contributing orbitals σ and π, with corresponding anti-bounding orbitals σ^* and π^*, respectively. Some chemical bounds present characteristic orbital conditions, ordered by higher to lower order energy transitions: i) alkanes ($\sigma \rightarrow \sigma^*$; 150nm); ii) carbonyls ($\sigma \rightarrow \pi^*$; 170nm); iii) unsaturated compounds ($\pi \rightarrow \pi^*$; 180nm); iv) molecular bounds to O, N, S and halogens ($n \rightarrow \sigma^*$; 190nm); and v) carbonyls ($n \rightarrow \pi^*$; 300nm). As most UV-VIS spectrometers yield a minimum wavelength of 200nm, this technique has been considered to provide lower information in terms of functional groups when compared to IR, being the spectral differences mostly attributed to conjugated $\pi \rightarrow \pi^*$ transitions and $n \rightarrow \pi^*$ transitions (Perkauparus et al., 1994).

Only recording $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions above the 200nm is however not totally a handicap. Many organic molecules present conjugated unsaturated and carbonyls bounds, such as aminoacids, phospholipids, free fatty acids, phenols and flavonoids, peroxides, peptides and proteins, sugars and their polymers absorbance in these bands. Furthermore, many biological molecules present chromophore groups, which increase the absortion in the UV-VIS region, such as: nitro, nitroso, azo, azo-amino, azoxy, carbonyl and thiocarbonyl, which can be used to identify microorganisms.

UV-VIS-SWNIR has some advantages to FTIR for microbiological identification in plate count agar. The lower wavelength turns this radiation attractive due to the lower penetration, being easier to monitor surfaces than NIR or MIR radiation. Furthermore, state-of-the-art fiber-optics miniature UV-VIS-SWNIR are today affordable for mobile applications such as identification of microorganisms in surfaces, using spectroscopy may became feasible in a near future. Although UV-VIS retrieves only molecular spectroscopy information, today's equipments also include high frequency vibrational spectroscopy in the SWNIR region, giving important information on water, fats and proteins which may be used to discriminate between microorganisms.

In this research was tried to discriminate both Yeasts and Bacteria of comonly used in microbiology laboratories: i) yeasts: *Saccharomyces cerevisiae*, *Saccharomyces bayanus*, *Candida albicans*, *Yarrowia lipolytica*, ii) bacteria: *Micrococcus luteus*,

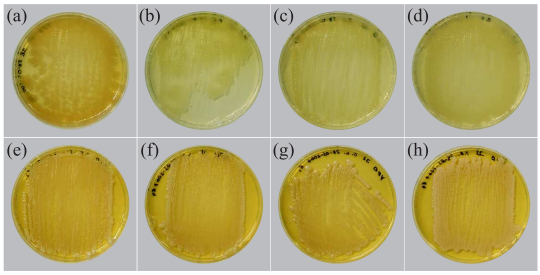


Figure 1: Yeast and bacteria growth media: (a) Tryptic Soy Agar (TSA): *Micrococcus luteus*; (b) Pseudomonas Isolation Agar (MP) (*Pseudomonas fluorescens*); (c) Luria-Bertani (LB): *Escherichia coli*, (d) LB: *Bacillus cereus*; (e) YPD (*Saccharomyces cerevisiae*); (f) YPD (*Saccharomyces bayanus*); (g) YPD (*Candida albicans*); and (h) YPD (*Yarrowia lipolytica*).

Pseudomonas fluorescens, *Escherichia coli*, *Bacillus cereus*; under plate count agar growth media.

The physical properties of the UV-VIS-SWNIR spectra can also provide a great potential for the identification of microorganism, using multivariate statistical analysis and signal processing techniques. Microbes may not be directly identified by their main colony chemical composition but rather by characteristic metabolites produced under different growth media. This is especially true for yeasts that exhibit one of the most complex metabolisms in this study. Therefore, not only the colony but changes in the composition of the plate count agar in the surroundings of each colony are expected to affect the UV-VIS-SWNIR spectra in order to obtain significant discrimination between the different microorganisms spectra. Therefore, the main objective of this research work were to investigate the discrimination potential of UV-VIS and VIS-NIR wavelengths to classify the following microorganisms: i) yeasts: *Saccharomyces cerevisiae*, *Saccharomyces bayanus*, *Candida albicans*, *Yarrowia lipolytica*, ii) bacteria: *Micrococcus luteus*, *Pseudomonas fluorescens*, *Escherichia coli*, *Bacillus cereus*; under plate count agar growth media.

2 MATERIALS AND METHODS

2.1 Sample Preparation

The microorganisms were obtained from the microbiological collection of the IBB - Institute for Biotechnology and Bioengineering at the University of Minho. The microorganisms were incubated under aerobic conditions at 35^oC during 72h. *Micrococcus luteus* was cultivated on Difco Tryptic Soy Agar (TSA) while Pseudomonas fluorescens was grown on Difco Pseudomonas Isolation Agar (MP). Escherichia coli and Bacillus cereus were cultivated on Difco Luria-Bertani Agar (LB). Yeast strains were grown on Difco YPD Agar (YPD), at the same temperature and time (Difco, 2005).

2.2 Spectroscopy

Yeast and bacteria UV-VIS-SWNIR spectroscopy analysis was performed using the fiber optic spectrometer AvaSpec-2048-4-DT (2048 pixel, 200-1100nm). Standart reflection UV-VIS and VIS-SWNIR probes, models FCR-7UV200-2ME and FCR-7IR200-2-ME (Avantes, 2007). A xenon and halogen light sources, models AvaLight XE-2000 and AvaLight-Hal were used for UV-VIS and VIR-SWNIR transmission measurements respectively; and recorded using AvaSoft 6.0 (Avantes, 2007). Transmission measurements were performed at the room temperature of 18 ± 2^oC, and: (a) UV-VIS: the xenon lamp was let to stabilize during 20 min; (b) VIS-NIR: the tungsten lamp lamp was let to stabilize during 15 min. The dark spectra was recorded and measurements were taken with linear and electric dark correction. Both light spectra were monitored by statistically assessing the reproducibility of the light source with measurements of light during the several days of the experiment. Fifteen spectra replicates were recorded of UV-VIS and VIS-SWNIR measurement of both plate count agar and microorganisms colonies to study scattering effects. Futhermore, spectra were obtained inside a box designed to isolate the environmental light and maintain the probe at 90^o angle with the plate agar.

2.3 Spectral Analysis

2.3.1 Spectra Pre-processing

Table 1 presents the UV-VIS and VIS-NIR spectra acquisition conditions. Experimental setup has shown that it is impossible to use the same integration time for the different microorganisms. Under these circumstances, all the collected spectra were normalized $\mathbf{x_{norm}}$ to remove this effect:

$$\mathbf{x_{norm}}_i = D_S \times \frac{\mathbf{x_{raw}}_i}{max(\mathbf{x}_i)} \quad (1)$$

where $\mathbf{x_{raw}}_i$ is the original spectra, D_S is the detector saturation value (14000 counts) and $\mathbf{x_{norm}}_i$ the i'th spectra normalized by its maximum value and resized to the detector saturation.

Furthermore, as most plaque count agar growth media are translucid, the signal recorded is in majority the media information. To increase spectral vari-

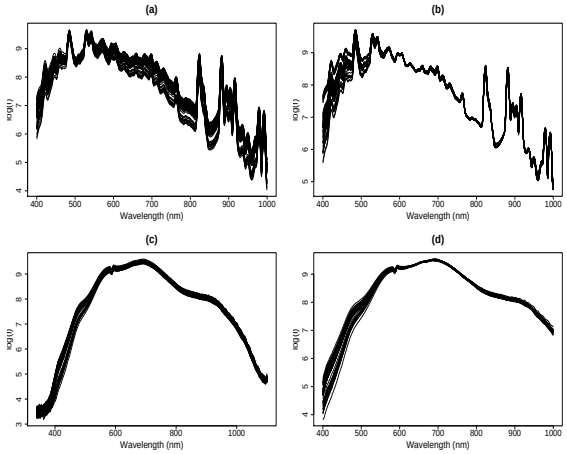


Figure 2: Plaque count agar spectra: (a) raw UV-VIS; (b) MSC UV-VIS; (c) raw VIS-NIR; (d) MSC VIS-NIR.

ance, the normalized media spectra matrix was subtracted to the microorganisms spectra, obtaining the spectra matrix (**x**), which is thereafter subjected to robust mean scattering correction, and singular value decomposition.

2.3.2 Robust Mean Scattering Correction

The collected spectra were smoothed by using a Savisky-Golay filter (length = 4, Order= 2) (**?**) prior to any exploratory data analysis procedure. Afterwards, the spectra was pre-processed using a modified multiplicative scatter correction algorithm (Gallagher et al., 2005; Martens and Stark, 1991; Martens et al., 2003). Each spectra is corrected by using the following equation:

$$\mathbf{x}_{corr} = \mathbf{x}b + a = \mathbf{x}_{ref} \tag{2}$$

The a and b are computed by minimizing the following error:

$$\mathbf{e}_j = b\mathbf{x}_j + a - \mathbf{x}_{ref} \tag{3}$$

where the $\mathbf{x}_j$ is the j sample spectra and $\mathbf{x}_{ref}$ is a reference spectra.

This RMSC algorithm is based on the application of the robust least squares method to determine the a and b matrices ensuring that spectral areas that do not correspond to scattering artifacts are not taken into account. The robust least squares algorithm is implemented by the re-weighted least squares with the weights computed by using the Huber function. The algorithm high breakdown point (50%) means that existent outliers will not distort the model fitting (eq. 3) and thus, the a and b scatter correction parameters are determined using only consistent spectral areas. The iterative algorithm can be described, briefly as follow: 1) set the reference spectra ($\mathbf{x}_{ref}$) equal to the sample

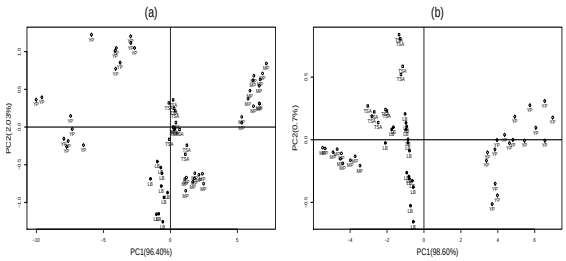


Figure 3: Growth media Gabriel plot: (a) UV-VIS (PC1 (96.40%), PC2 (2.3%) and (b) VIS-NIR (PC1 (98.60%), PC2 (0.70%) to the corresponding growth media: TSA (□), MP (○), LB (●) and YPD (◇).

spectra closest to the median spectra; 2) correct the remaining sample spectra by applying the above described robust least squares procedure; eq.3) recompute the median spectra and iterate until convergence.

2.3.3 Singular Value Decompostion

Singular value decomposition (SVD) is a blind signal technique widely used in spectroscopy data, where the corrected spectra ($\mathbf{x}_{corr}$) is decomposed in order of magnitude of variation directions in the variable space (wavelengths). Generally, most variability is captured in the first principal components (PC), where as, in good signal to noise spectral data, noise is captured in the last orthogonal decompositions, and therefore a spectra can be decomposed as:

$$\mathbf{x}_{corr} = \widehat{\mathbf{x}} + \varepsilon(\mathbf{x}) \tag{4}$$

where the $\widehat{\mathbf{x}}$ is the signal and $\varepsilon(\mathbf{x})$ is the estimated noise of $\mathbf{x}$. Spectra matrix $\mathbf{x}_{corr}$ can be decomposed by (SVD), where:

$$\mathbf{x} = \mathbf{USV^T} \tag{5}$$

where **US** are the scores, $\mathbf{V^T}$ the loadings and the **S** singular values, respectively (Jolliffe, 1986; Krzanowski, 1998; Baig and Rehman, 2006).

To distinguish between the number of relevant decompositions, one can determine the number of relevant singular values by performing n randomizations of the original spectra matrix (**x**) (Manly, 1998). In this research, 5000 randomizations were performed by rotating the spectral scope value at the same wavelengths among the different samples, in order to not violate the spectral continuity. Singular values from the original spectra **x** above the 1st singular value of randomized spectra ($\mathbf{x}_{rand}$) define the number of independent singular values of the original signal where is possible to discriminate the different microorganisms spectra:

$$\widehat{\mathbf{x}} = \mathbf{US}_{relv}\mathbf{V^T}_{relv} \tag{6}$$

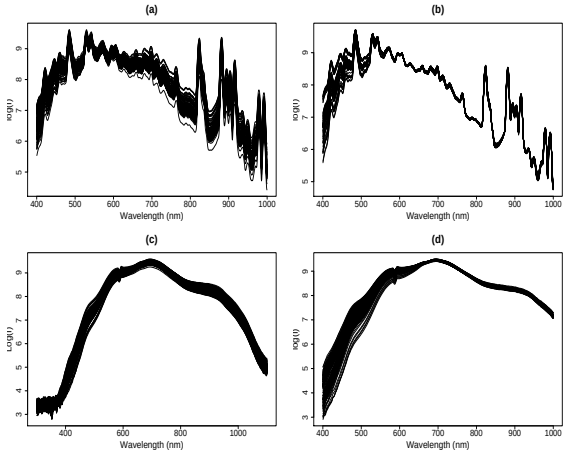


Figure 4: Microorganisms spectra: (a) raw UV-VIS; (b) MSC UV-VIS; (c) raw VIS-NIR; (d) MSC VIS-NIR.

Where $\mathbf{US_{relv}}$ and $\mathbf{V^T_{relv}}$ are the statistically relevant scores and loading of $\mathbf{x}$, respectively. To further discriminate between the microorganisms spectra, the relevant PC's scores($\mathbf{US}_{relv}$) were subjected to hierarchical clustering according to the euclidean distance, to determine the potential of using UV-VIS and VIS-NIR to identify the studied yeast and bacteria. All statistical computing analysis were performed using R (R-Project, 2006).

3 RESULTS AND DISCUSSION

3.1 Spectral Analysis

Figure 2 presents the UV-VIS and VIS-NIR plate count agar growth media spectra, respectively. It is possible to observe in Figure 2(a) and 2(c) that all plate count agar are highly dispersive, generating a high scattering effect. This undesired scattering effect is due to the light path length to be very sensitive to the probe angle, particles in the agar, surface texture of both agar and petri disk. If the light scattering effect is not corrected, variance due to this physical phenomenon affects significantly the chemical interpretation of the spectra due to scattering artifacts.

As pure additive effects of light scattering are rarely observed in samples with complex compositions, being mainly of multiplicative origin, the growth media spectra was subjected to RMSC, being the corrected spectra presented in Figures 2 (b) and 2 (c), respectively. By directly comparing Figures 2 (a)-(b), and 2 (c)-(d), one can observe the scattering effect is obtained in both light sources. After applying the RMSC one can observe that, this scattering artifacts are significantly reduced in the region of 700-1000nm, but nevertheless both light sources present higher degree of spectral variance in the region of 400-700nm.

Furthermore, it is observable that all spectra are proportional to each other. Variation is mostly in terms of signal intensity than in spectra shape. In this sense it is difficult to distinguish the different growth media by direct spectra comparison. Figure 3 presents the Gabriel plot (PC1 vs PC2) of the growth media spectra, for UV-VIS and VIS-NIR wavelengths, respectively. Both UV-VIS and VIS-NIR biplots evidence optical properties of the growth media, being thin mostly described by the 1st principal component. In both biplots, it is possible to observe that the MP growth media is the most translucid and YPD the most opaque, and therefore PC1 can be interpreted as the amount of signal that the detector records. Media such as TSA and LB present similar spectra records. Such is mainly attributed to their similar composition in terms of main components such as sodium chloride, agar and water.

Results show that MP and YPD media are more suceptible to variability than TSA and LB in spectroscopy terms. As no chemical assessment was performed to the media, we cannot present the cause for this source of variation.

Figure 4 presents the UV-VIS and VIS-NIR microorganisms spectra, respectively. Similarly to the growth media, the microbe spectra exhibits high scattering artifacts (see Figures 4a and 4c). The scattering effect is in this case due to the light scattering at the colony surface and growth media, which significantly affects the observed spectra. Similarly to the growth media, scattering is significantly high in both light sources and evenly distributed from 450 to 1000 nm, nevetheless scattering is significantly reduced by the RMSC algorithm. The corrected spectra presents higher variability in the region of 400-700nm, but nevertheless it is difficult to recognize directly spectral characteristics that distinguish the different microorganisms under study, and therefore SVD analysis is necessary.

3.2 Singular Value Decomposition Analysis

Figure 5 presents the Gabriel plot of the first two components obtained by SVD (PC1(78.40%), PC2(8.03%)), and the corresponding hierarchical clustering analysis. The first two decompositions ensure the majority of the spectral variation to proportional to the average spectrum (PC1(78.40%)), and that linear variance also with discriminant power (8.03%), evidences smaller but discriminate spectral

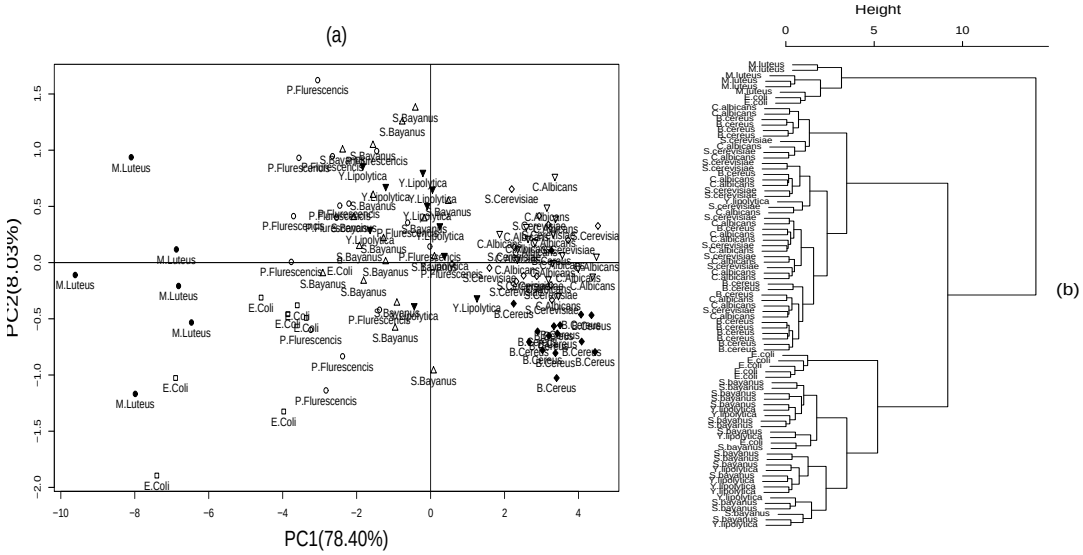


Figure 5: UV-VIS Microorganisms Spectra PCA Analysis: (a) Gabriel Plot (PC1 (78.40%), PC2 (8.03%); *S. cerevisiae* (◇), *S. bayanus* (△), *C. albicans* (▽), *Y. lipolytica* (▼), *M. luteus* (●), *P. fluorescens* (○), *E. coli* (□) and *B. cereus* (⧫); (b) hierarquical clustering of Microorganisms.

differences between the studied microorganisms.

PC1(78.40%) discriminates between spectral intensity, being possible to observe that colonies of *B. cereus*, *C. albicans* and *S.cerevisiae*, present the higher scores indicating that the colonies of this microorganisms are well suited for diffusive reflectance. Although the relative proximity in the scores space, it is observable that *B. cereus*, *C. albicans* and *S. cerevisiae* are discriminated by the 2nd PC. In the group, it is possible to observe a higher similarity between *C. albicans* and *S.cerevisiae* spectra than with *B. Cereus*.

It is further observable that although *S. bayanus* and *Y. lipolityca* exhibit similar spectral intensity, their spectra is possible to he discriminated in the 2nd PC. *S. bayanus* who presents larger variability then *Y. lipolityca*, being more difficult to identify.

E. coli and *M. luteus* colonies are distinguishable from all the other microorganisms group. These present the lowest signal intensity, and the proximity of the two spectra may in part he due to the growth media spectral similarity between TSA and LB, as show in Figure 3a. Nevertheless, the UV-VIS spectra decomposition is capable of discriminating between *E. coli* and *M. luteus* spectra.

In the UV-VIS, *P. fluorescens* spectra has proved to he highly unreproducible, being its scores well spread throughout the 2nd PC. This unreproducibility is attributed to the high translucent of both *P. fluorescens* colonies and the MP growth media, and to the experimental microbiological technique. As microorganisms were inoculated using a inoculating loop, a significant part of radiation is diffused into de media in *P. fluorescens*. Better growth may in the future improve the spectral measurement and therefore its identification, as already observed in *E. coli* colonies, which although are smaller, are capable of developing thicker colonies.

Dispite the experimental difficulties hierarchical clustering analysis presented in Figure 5b shows that the majority of the studied microorganisms cluster together with exception of *P. fluorescens*. This gives good perspectives of using UV-VIS spectroscopy for microorganisms identification in plate count agar after experimental and signal processing improvements.

The VIS-SWNIR spectra also exhibits high scattering artifacts in the 400-1000nm region (see Figure 4c), the scattering effect is effectively removed by the RMSC algorithm used during the spectral preprocessing. Figure 4d, shows that in the corrected spectra, variance is higher in the 400-700nm as already observed as with the UV-VIS light source. The majority of the spectra are proportional to each other, varying only in signal intensity.

Figure 6a presents the Gabriel plot of the first two PCs of the decompose VIS-NIR spectra (PC1 (75.40%), PC2 (10.95%), and the corresponding scores hierarchical clustering. Similarly to the UV-VIS spectra, the first spectra decomposition discriminating variance proportional to the average spectra, and PC2 (lower variance) tends to discriminate smaller details between the spectra of the studied microorganisms, evidencing that VIS-NIR is a viable methodology for identification microorganisms on plate count agar. PC1 (75.40%) of VIS-SWNIR spec-

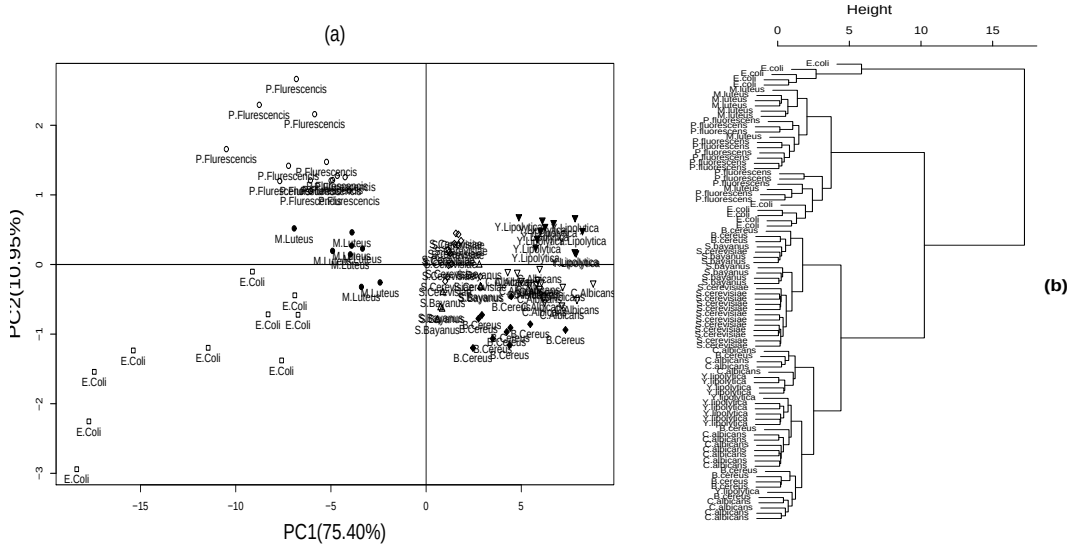


Figure 6: VIS-NIR Microorganisms SpectraPCA Analysis: (a) Gabriel Plot (PC1 (75.40%), PC2 (10.95%); *S. cerevisiae.* (⋄), *S. bayanus* (△), *C. albicans* (▽), *Y. lipolytica* (▼), *M. luteus* (●), *P. fluorescens* (○), *E. coli* (□) and *B. cereus* (⧫); (b) hierarquical clustering of Microorganisms.

tra presents a clear discrimination between groups of microorganisms: (a) *B. cereus*, *S.cerevisiae*, *S. bayanus*, *Y. lipolityca*, *C. albicans*, and (b) *M. luteus*, *E. coli* and *P. fluorescens*. PC2 (10.95%) is capable of discriminate the microorganisms inside these two groups, and therefore, complete discrimination is achievable with VIS-SWNIR wavelengths.

In the VIS-SWNIR spectra *E. coli* and *P. fluorescens* present higher dispersion of signal when compared to the rest of the spectra of microorganisms. Nevertheless, results reproductivity inside these two groups is significantly higher than with the UV-VIS measurements; indicating that translucid colonies were better identified under the VIS-SWNIR radiation. Furthermore, data suggest that VIS-SWNIR may be better to be used under non-optimal measuring conditions. This is especially problematic if we want to identify microorganisms in plate count agar with similar compositions and metabolism.

3.3 Methodology Improvements

This preliminary study shows that UV-VIS and VIS-SWNIR reflection spectroscopy has a great potential for rapid qualitative discrimination of yeasts and bacteria in plate count agar. Nevertheless, both experimental methodology and signal processing techniques should be improved to take advantage of the information contained in the UV-VIS-SWNIR.

Improvements to the experimental methodology are necessary to improve discrimination in the studied region of the spectra. For example, the use of liquid cultures to replicate microorganisms and then innoculate by a droplet on the surface agar will allow the growth of compact and thicker colonies. Small and thin colonies lead to readings with much media information, because the light is reflected from both sites of the agar. This is particularly relevant for *E. coli* and *P. fluorescens* because of the small size of the colonies and small thickness of *P. fluorescens*; being the spectra highly affected by the media composition, leading to sistematic errors in discrimination. Microorganisms signal spectra can also be maximised by reducing the growth media thickness. Such minimises light dispersion accross the agar, and increases the colony spectral intensity by passing through it the reflected light; ans as well by optimising fiber optics diffusive reflectance position and angle control, minimizing scattering effects in the colonies and growth media.

Scattering artifacts and small noise were successfully removed by pre-processing the spectra with the RMSC and Savisky-Golay filter, respectively; being possible to achieve high-quality and resolution final spectra before signal treatment. Improvements can also be performed to the spectra processing procedure. Methods such as the combination of logistic partial least squares (log-PLS) with multiblock (e.g. UV-Vis+Vis-Nir spectra) can provide a new insight to the discrimination of microorganisms using spectroscopy. If these methodologies provide good discriminations, more robust techniques such as the use of wavelets for compressing the original spectra and orthogonal-PLS classification of spectra will be tested.

4 CONCLUSIONS

Results show that UV-VIS-SWNIR spectroscopy is a feasable technology for plate count agar microorganisms identifications. The robust mean scattering correction algorithm was able to efficiently remove the growth media and colonies scattering artifacts, allowing a better interpretation of the singular value decomposition scores loading. In this exploratory experiment, VIS-SWNIR wavelengths were able to produce better discriminations between microorganisms than the UV-VIS region. Nevertheless, experimental methodology and signal processing improvements proposed may increase the discrimination resolution, making UV-VIS-SWNIR an attractive methodology for rapid microorganisms identification in plate count agar.

REFERENCES

Avantes (2007). Avaspec-2048-4-dt/-rm. http://www.avantes.com.

Baig, S. and Rehman, F. (2006). Signal modeling using singular value decomposition. In *Advances in Computer, Information, and Systems Sciences, and Engineering*. Springer Netherlands.

Berger, A. and Zhu, Q. (2003). Identification of oral bacteria by raman microspectroscopy. In *Journal of Modern Optics 50(15-17): 2375-2380*. Taylor and Francis Group.

Bhatta, H., Goldys, E., and Learmonth, R. (2005). Rapid identification of microorganisms by intrinsicfluorescence. In *Imaging, Manipulation, and Analysis of Biomolecules and Cells: Fundamentals and Applications III,*. SPIE.

Difco, L. (2005). *Difco manual : dehydrated culture media and reagents for microbiology. - VII*. Difco Laboratories, Detroit, USA.

Dziuba, B., Babuchowski, A., Naleczb, D., and Niklewicz, M. (2007). Identification of lactic acid bacteria using ftir spectroscopy and cluster analysis. In *International Dairy Journal 17: 183189*. Elsevier.

Gallagher, N. B., Blake, T., and Gassman, P. (2005). Application of extended inverse scattering correction to mid-infrared reflectance of soil. In *Journal of Chemometrics 19: 271-281*.

Guibeta, F., Marieya, L., Pichonb, P., Traverta, J., Denisb, C., and Amiela, C. Discrimination and classification of enterococci by fourier transform infrared spectroscopy.

Hammes, G. G. (2005). Fundamentals of spectroscopy. In *Spectroscopy for the Biological Sciences*. John Wiley Sons, Inc.

Harz, M., Rosch, P., Peschke, K.-D., Ronneberger, O., Burkhardtb, H., and Popp, J. (2005). Micro-raman spectroscopic identification of bacterial cells of the genus staphylococcus and dependence on their cultivation conditions. In *Analyst 130: 15431550*. The Royal Society of Chemistry.

Jolliffe, I. (1986). *Principal Component Analysis*. Springer, New York, USA.

Krzanowski, W. J. (1998). *Principles of Multivariate Data Analysis*. Oxford University Press, Oxford, UK.

Manly, B. F. (1998). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall, London, UK, 2nd edition.

Mariey, L., Signolle, J., Amiel, C., and Travert, J. (2001). Discrimination, classification, identification of microorganisms using ftir spectroscopy and chemometrics. In *Vibrational Spectroscopy 26: 151-159*. Elsevier.

Martens, H., Nielsen, J. P., and Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. In *Analytical Chemistry 75(9): 394-404*. American Chemical Society.

Martens, H. and Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. In *Journal of Pharmaceutical and Biomedical Analysis 9: 625-635*. American Chemical Society.

Oberreuter, H., Mertens, F., Seiler, H., and Scherer, S. (2000). Quantification of micro-organisms in binary mixed populations by fourier transform infrared (ft-ir) spectroscopy. In *Letters in Applied Microbiology 2000, 30: 8589*. The Society for Applied Microbiology.

Oberreuter, H., Seiler, H., and Scherer, S. (2002). Identification of coryneform bacteria and related taxa by fourier-transform infrared (ft-ir) spectroscopy. In *International Journal of Systematic and Evolutionary Microbiology 52: 91100*. IUMS.

Perkauparus, H., Grinter, H., and Therfall, T. (1994). *Uv-Vis spectroscopy and its applications*. Springer-Verlag, New York, USA.

R-Project (2006). R: A programming environment for data analysis and graphics. URL: http://www.r-project.org.

Rch, P., Harz, M., Schmitt, M., and Popp, J. (2005). Raman spectroscopic identification of single yeast cells. In *J. Raman Spectrosc 36: 377379*. Wiley InterScience.

Riddle, J., Kabler, P., Kenner, B., Bordner, R., Rockwood, S., and Stevenson, H. (1956). *Bacterial Identification by Infrared Spectrophotometry*. Cincinnati, Ohio.

Winder, C., Carr, E., Goodacre, R., and Seviour, R. (2004). The rapid identification of acinetobacter species using fourier transform infrared spectroscopy. In *Journal of Applied Microbiology 96, 328339*. The Society for Applied Microbiology.

FUZZY MRF MODELS WITH MULTIFRACTAL ANALYSIS FOR MRI BRAIN TISSUE CLASSIFICATION

Liang Geng and Weibei Dou
Department of Electronic Engineering, Tsinghua University, Beijing, P. R. China
douwb@tsinghua.edu.cn

Keywords: Brain tissue classification, Fuzzy MRF Model, multifractal analysis.

Abstract: This paper introduces multifractal analysis to the Fuzzy Markov Random Field (MRF) Model, used for brain tissue classification of Magnetic Resonance Images (MRI). The traditional classifying method using Fuzzy MRF Model is already able to calculate out the memberships of each voxel, to solve the Partial Volume Effect (PVE). But its accuracy is relatively low, for its spatial resolution is not high enough. Therefore the multifractal analysis is brought in to raise the accuracy by providing local information. The improved method is tests on both simulated data and real images, where results on membership average errors and position errors are calculated. These results show that the improved method can provide much higher accuracy.

1 INTRODUCTION

Magnetic Resonance Images (MRI) have been widely used for brain diagnosis and disorder detections. Accordingly, segmenting brain images into different tissues, such as cerebrospinal fluid (CSF), grey matter (GM) and white matter (WM), for clinical uses, has become a classical problem.

Many different tissue segmenting methods and algorithms are proposed these years. Some methods are using T1 weighted images (Rajapakse *et al*, 1996), while others use multispectral MR data (Taxt and Lundervold, 1994). Algorithms can be based on histogram determination Suzuki and Toriwaki, 1991), or on a priori information on anatomy (Joliot and Mazoyer, 1993). Mathematical models are used, from cluster analysis (Simmons *et al*, 1994) to Bayesian estimation (Chang *et al*, 1996). All these methods assume that each voxel in the images to be segmented belongs to only one specific tissue. However, due to the partial volume effect (PVE), one voxel may contain information from several different tissues, flawing the segmenting results of the methods proposed.

To solve the effect of PVE, Markov Random Field (MRF) Model is applied to tissue classification (Ruan and Cyril *et al*, 2000). The a priori information from an image and the classifying criteria are combined into energy functions of MRF's distribution, and then the voxels with mixed tissues can be classified by the iterated conditional mode (ICM). This method achieves a so-called 'Hard Classifying', classifying each voxel into one tissue who contributes the most, and contributions from other tissues are neglected. Considering that the neglected information is usually useful, a further model, the Fuzzy MRF Model, is brought in (Ruan and Moretti *et al*, 2001). The Fuzzy MRF Model takes into account the contextual information, the statistical information and the anatomical information of the brain. And 'Hard Classifying' is replaced by 'Fuzzy Classifying', providing 'memberships' for each voxel, indicating each voxel's partial volume degree, in other words, representing how much these tissues occupy one voxel respectively.

The fuzzy MRF Model is proved effective on PVE, but still limitations it has. Experiments show that this method performs poorly at brinks of brain images, where grey-level of voxels changes suddenly, which implies its spatial resolution is not high enough. Also, this method being noise sensitive, when it encounters images with high noise, its accuracy becomes even worse. These limitations can be attributed to the lack of local properties extracted from images, so what we need to do is to provide the Fuzzy MRF abundant local information.

As a new signal processing method, multifractal analysis is competent for this object. Multifractal is

first studied mathematically (Halsey *et al*, 1986), and introduced to image processing by Sarkar and Katsuragawa (1995). It has derived various methods for image analysis, and has shown its advantages in local feature extraction (Liu and Li, 1997). It is also adapted to MRI brain tissue classifying, to remove ambiguities in the 'Hard Classifying' caused by intensity overlap, and performed well (Ruan *et al*, 2000)

Our research aims to raise the spatial resolution by local information while using fuzzy MRF model. We propose a combining both fuzzy MRF model and multifractal analysis together, to achieve a more accurate 'Fuzzy Classifying'. In this paper, we firstly show an overall of the proposed scheme and two kenel algorithms, fuzzy MRF and multifractal analysis, then explain how to combine these two parts in section 2. The validation of this improved scheme is done both by some experiments and in comparison with traditional fuzzy MRF method. The results and discussion are shown in section 3. This improved algorithm takes the same frame as the original method, while changes are done mathematically. Experiments and tests are done on various images, including real and virtual data with different amount of added noise.

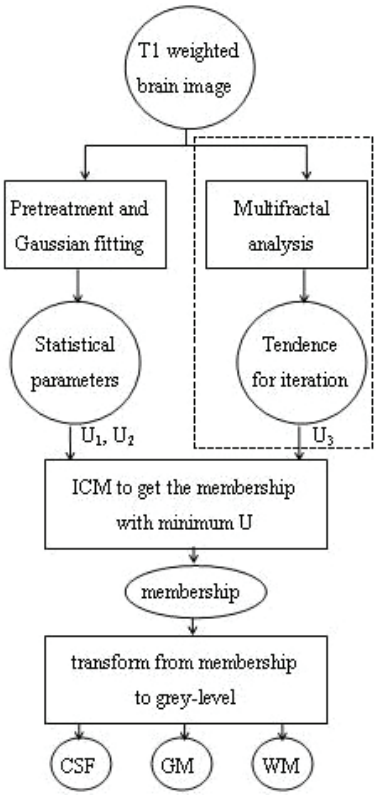


Figure 1: Flowchart of the whole algorithm.

2 ALGORITHMS

In this section we will introduce the algorithms for Fuzzy MRF Model along with multifractal analysis. We will show how the Fuzzy MRF Model works and how the multifractal information improves its classifying results.

2.1 A Whole Algorithm for Fuzzy MRF Model with Multifractal Analysis

Here we give out the flowchart of the whole algorithm using Fuzzy MRF Model with multifractal analysis, as Figure 1.

A parallel treatment, such as a preteatment for the parameter estimation of Fuzzy MRF Model and a multifractal analysis for producing a novel parameter U_3 to adjust a traditional Fuzzy MRF Model, is the contribution of this scheme.

The region framed by dashed line rectangle is the multifractal part added to the original frame. The other modules form the ICM iteration of Fuzzy MRF (Ruan and Moretti *et al*, 2001), and the multifractal part provides a 'tendence' to instruct the iterating course. We will discuss them in detail in following subsections.

2.2 Fuzzy MRF Model

The MRF Model is an a priori model, it represents the spatial correlation of image data. Considering a random field A with its realization a, in practice we usually use the joint probability density function of A on the whole image. Particularly when the probability density is distributed Gibbsian, the density function takes form as (1):

$$P(X=x)=\frac{1}{Z}\exp(-U(x)),\qquad Z=\sum_{x}\exp(-U(x)) \tag{1}$$

where $U(a)$ stands for the energy function, and Z the normalizing constant.

Fuzzy MRF Model applied to image segmentation, there are two random fields. One is the membership field A, whose realization is a, the other is the grey-level field Y, whose realization is y, which is known a priori. The goal of tissue classifying is to achieve the maximum joint probability density distribution of these two random field:

$$a_{result} = a \mid \{P_{A|Y}(a,y) \geq P_{A|Y}(x,y),\ \forall x\} \qquad (2)$$

The joint probability can be represented by conditional probability as:

$$P_{A,Y}(a,y) = P_A(a)P_{Y|A}(y \mid a) \qquad (3)$$

Comparing (1) and (3), we can get the probability distribution of Fuzzy MRF of image:

$$P_{A,Y}(a,y) = \frac{1}{Z}\exp(-U_1(a,y) - U_2(a)) \qquad (4)$$

Here U_1 represents the incompatibility between the grey-levels and the memberships, and U_2 represents the inhomogeneity of memberships themselves. They can be calculated using statistical parameters, which are acquired by fitting the grey-level histogram with several Gaussian functions (Ruan and Jaggi *et al*, 2000).

Once the two parts of energy function are calculated out, we can use the deterministic relaxation iterated conditional modes (ICM) to find the optimum realization of membership a, to ensure the energy function U being minimum, which means the joint probability in (1) being maximum.

The original algorithm concerns only these two parts of energy function, and information about the partial details are not taken into account. So we can see the shortcome of the original algorithms clearly by calculating the set-difference between classifying resluts and standard modules. Here we use a noise-free virtual image of normal brain with no RF. The original image is shown in Figure 2. Classifying results are shown in Figure 3 and differences in Figure 4, as we can see, the spatial differences mainly locate on the brinks, stings and nicks of the image, where grey level changes suddenly. If we could provide the algorithm enough local information to raise its spatial resolution, the result should be more accurate.

2.3 Multifractal Analysis

The multifractal analysis is first adopted into 'Hard Classification' by Ruan (Ruan and Bloyet, 2000), to remove the ambiguity caused by intensity overlap. The intensity overlap has nothing to do with the fuzzy model, since in fuzzy circumstances, we need not to reclassify a mixed voxel into one particular pure tissue. But the local information provided by multifractal still helps in raising the spatial resolution, thus we introduce the multifractal method to the Fuzzy MRF Model.

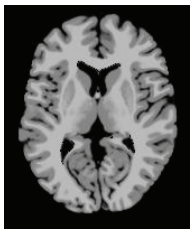


Figure 2: The original image named Vn00.

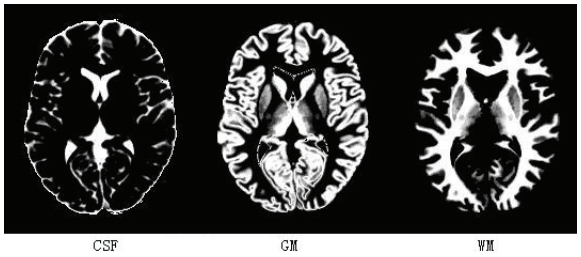


Figure 3: The classifying results of a virtual image.

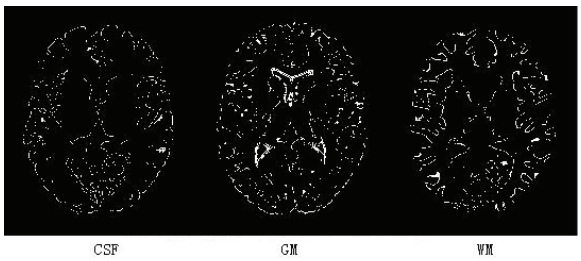


Figure 4: The spatial difference between classifying results and standard modules of a virtual image.

2.3.1 Multifractal in Signals

It is well known that fractal is widely used to process self-similar signals, by providing its global information of similarity to the 'fractal dimension'. But to provide local information, we need the 'fractal dimension' to vary from part to part of the signal. This is multifractal.

Therefore Multifractal dimension is defined locally by the measurement and length of a shrinking small region, as (5):

$$\alpha = \lim_{a \to 0} \frac{\log b}{\log a} \qquad (5)$$

where α denotes the multifractal dimension, also called Hölder exponent, b denotes the measurement, and a the length of the region.

Each small region has its own Hölder exponent, and then the whole signal can be considered as the union of many subsets that combining with each other. To characterize the local characteristics, we need another parameter to decompose these small regions, and group all voxels being in the same kind of detail into a set. The parameter brought in is called 'multifractal spectrum', defined as $f(\alpha)$. $f(\alpha)$'s definition can be Hausdorff, Legendrea, or others. We can also define it particularly.

2.3.2 Multifractal in Brain Images

To describe the local details of brain images, first we need to abstract these details into several simple models. Observe the images, we can find out three kinds of details shown in Figure 5.

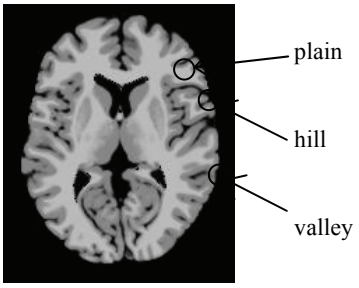
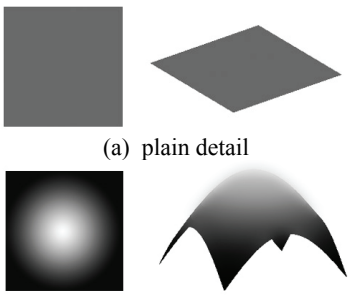


Figure 5: The details of brain image.



(a) plain detail

(b) hill detail

(c) valley detail

Figure 6: The models of details. The intensity model is on the left, while grey-level model is on the right.

Grey levels of voxels in plain region has little difference from the central voxel, most of the small regions are proved to be plain. Hill region has several voxels much lighter in the centre, and valley region has a much darker centre. The models can be illustrated as Figure 6.

After defined the three detail models, the Hölder exponent α is ready to be calculated out for each model. From the equation (5) we could know that α is defined to be a limit process. Because the image is composed by discrete voxels, the values of length *a* must also be discrete, thus the limit process is discrete: first *a* takes the radius of the small region *R* as its value, then each time *a* minus 1 until *a* becomes 0. The corresponding value of *b* is the sum of grey level of voxels in a diminishing spherical small region whose radius is *a*. Both *a* and *b* gotten, the Hölder exponent α can be gotten in succession. Since we only care about the relative size of the Hölder exponent α, the values themselves make no sense to us; we can also use some approximate method, such as linear fitting, instead of the complicate limit process.

At last, we can get the relative size of the Hölder exponent α in different details: for hill, α is relatively smaller, and for valley, α is relatively bigger, while for plain, it's in the middle.

To decompose image details and group the voxels into three sets, $f(\alpha)$ needs to be generated from α. And for concision, we define $f(\alpha)$ as α's histogram, that means:

$$f(\alpha_i) = \sum_{k \in I} \delta(\alpha(k), \alpha_i) \qquad (6)$$

where *I* represents the whole image, $\alpha(k)$ is the Hölder exponent at voxel *k*, $\delta(\alpha(k), \alpha_i)$ is Kronecker Function, which takes the value 1 while $\alpha(k) = \alpha_i$, and 0 while $\alpha(k) \neq \alpha_i$.

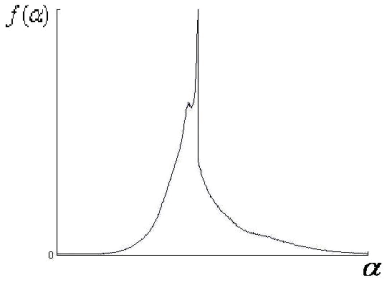


Figure 7: Multifractal spectrum of MR brain image.

Then we get the histogram as spectrum, shown in Figure 7. A correctly collected MR Brain image must has a multiractal spectrum in this shape, because most voxels are in plain regions, which makes the high peak in the middle. Therefore we need only to find the position of the peak, denoted as

α_0, representing the corresponding voxels being in plain detail, and voxels with an α smaller than α_0 are in hill regions, others are in valley regions.

2.4 Multifractal Applied to Fuzzy MRF

Using multifractal, we can label every voxel the detail type it belongs to, and the rest to be done is to combine the multifractal and Fuzzy MRF together, by influencing the ICM iterating process with these detail labels.

We consider translating the labels into some sort of 'tendence'. If a voxel is labelled 'hill', that means it's brighter than its neighbours, then it should have a tendence to be classified into a brighter tissue. If the voxel is labelled 'valley', on contrary, it should have a tendence to be classified into a darker tissue. If the voxel is labelled 'plain', its brightness is almost the same as its neighbours', so it should have no tendence.

Then the main problem is how to translate the detail labels into 'tendences'. Here we propose the 3rd energy function U_3, to change the value of U, therefore to impose the 'tendence' to the iterating process. From to gradient label (denoted by D), then to U_3, can be defined as equation (7):

$$D=\begin{cases}1 & \alpha<\alpha_{hill} \quad (hill)\\ 0 & \alpha_{hill}\le\alpha\le\alpha_{valley} \ (plain)\\ -1 & \alpha>\alpha_{valley} \quad (valley)\end{cases},$$
$$U_3 = D\cdot\beta_{fractal}\cdot\begin{cases}-1 & a>a_{current}\\ 0 & a=a_{current}\\ 1 & a<a_{current}\end{cases},\ \beta_{fractal}>0, \tag{7}$$

For equation (7), α_{hill} and α_{valley} are thresholds generated from the spectrum $f(\alpha)$ shown in Figure 7, e.g, $\alpha_{hill}=\max(\alpha\mid f(\alpha)<f(\alpha_0)/2\ \&\ \alpha<\alpha_0)$ and $\alpha_{valley}=\min(\alpha\mid f(\alpha)<f(\alpha_0)/2\ \&\ \alpha>\alpha_0)$. And $\beta_{fractal}$ is a positive weight coefficient for U_3, whose value depends on how much you want the multifractal part to affect the whole system.

Using (7), the detail 'hill' can make U with brighter membership a smaller, and U with darker membership a bigger. For the detail 'valley', the performance is on the contrary. Thus multifractal can be applied to the algorithm frame shown in Figure 1.

3 EXPERIMENTS AND RESULTS

3.1 Experiment Materials

Experiments are done on 9 data sets to test the improved algorithm. These 9 sets of data includes various conditions, such as virtual data and real images, data with different noise levels and RF levels, data of normal brains and brains with defect. We name each image the way as following. The 1st letter indicates its source in V (virtual) and R (real). The 2nd letter indicates the defect of the brain, in n (normal), s (multiple sclerosis) and t (tumour). The 1st number indicates its noise level in percent. And the 2nd number indicates whether RF is added, in 1 if added or 0 if not.

The information of the 9 sets of data is listed in Table 1.

Table 1: Information of data sets used for tests.

Name	Source	Defect	Noise	RF
Vn00	virtual	normal	0%	0%
Vn30	virtual	normal	3%	0%
Vn50	virtual	normal	5%	0%
Vn70	virtual	normal	7%	0%
Vn01	virtual	normal	0%	20%
Vn71	virtual	normal	7%	20%
Vs00	virtual	multiple sclerosis	0%	0%
Rn	real	normal		
Rt	real	tumour		

To quantify the tests of accuracy, we mainly use the virtual data and their standard modules. The virtual data is from Montréal Neurological Institute, McGill University, McConnell Brain Imaging Centre (Website: http://www.bic.mni.mcgill.ca/brainweb/).

3.2 Evaluating Method

The classifying results of virtual images are evaluated in two ways. The 1st way is the position error e_p, which is the number of voxels classified differently from the standard module. The position error is defined as equation (8).

$$e_p = \sum_{(i,j,k)\in I} \delta(a_{result(i,j,k)}, a_{std(i,j,k)})$$

$$\delta(a_{result(i,j,k)}, a_{std(i,j,k)}) = \begin{cases} 1 & a_{result(i,j,k)} = a_{std(i,j,k)} \\ 0 & a_{result(i,j,k)} \neq a_{std(i,j,k)} \end{cases} \quad (8)$$

And the 2nd way is the membership average error e_m, which indicates the average error of memberships from the whole images. The membership average error is defined as equation (9), where N(I) represents the number of voxels.

$$e_m = \frac{\sum_{(i,j,k)\in I} \left| a_{result(i,j,k)} - a_{std(i,j,k)} \right|}{N(I)} \quad (9)$$

3.3 Result and Discussion

The position error of each image data using each algorithm is listed in Table 2, the membership average errors are listed in Table 3.

Both Table 2 and Table 3 show that the algorithm with multifractal has lower errors, in other words, higher accuracy than the original one. (In spite of some exceptions caused by noise and RF, such as GMs of Vn50 and Vn01 in Table 2, the flaws can be compensated by better results on the other tissues.)

Table 2: Position errors of two algorithms.

Data	Multi-fractal	e_p (number of voxels)		
		CSF	GM	WM
Vn00	without	68232	105738	55071
	with	66398	98904	54122
Vn30	without	114829	140443	139721
	with	115507	134358	139774
Vn50	without	188361	193855	223821
	with	183362	194035	222067
Vn70	without	228503	238072	281032
	with	228273	230507	278875
Vn01	without	165458	255996	195482
	with	166628	256531	190323
Vn71	without	232630	256904	306778
	with	232336	250765	302827
Vs00	without	72145	124560	72968
	with	72170	119322	69315

Table 3: Membership average errors of two algorithms.

Data	Multi-fractal	e_m		
		CSF	GM	WM
Vn00	without	1.6322	3.3685	1.3824
	with	1.5104	3.1752	1.3422
Vn30	without	2.1395	4.9949	2.7958
	with	2.0264	4.8018	2.7688
Vn50	without	2.9763	7.4779	4.6330
	with	2.8443	7.2597	4.5960
Vn70	without	4.3948	9.4826	5.9437
	with	3.5422	9.1410	5.9312
Vn01	without	2.3392	6.2700	3.9262
	with	2.2365	6.1273	3.8102
Vn71	without	3.8495	10.526	7.0626
	with	3.6663	10.136	7.0032
Vs00	without	1.6992	3.4159	1.4649
	with	1.6233	3.2771	1.4197

Because of the effect of other tissues such as muscles and bones, the errors are still not very low, but we could observe just the voxels at brinks, which we care about. Comparing the result images, we find that the voxels improved are mainly what we wanted to improve. Compare to the results from original method, the results of multifractal method have much less error voxels at the brinks of images. One comparison of position error using Vn00, the same data as Figure 2, is shown in Figure 8.

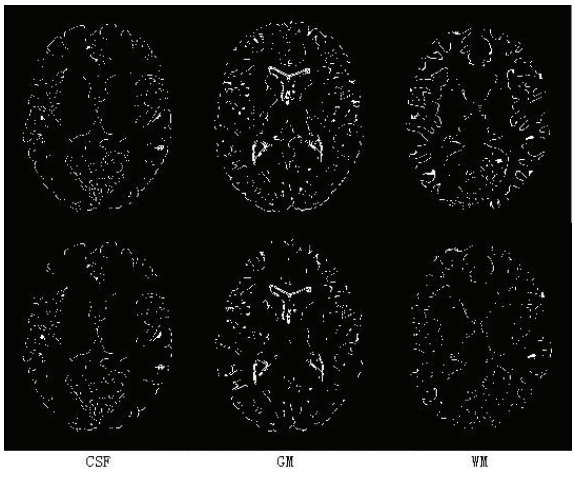


Figure 8: Position error of original algorithm (above) and improved algorithm with multifractal (below).

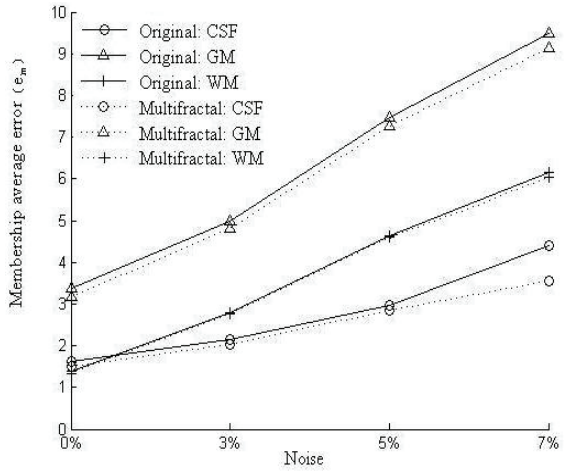


Figure 9: Membership average error of original algorithm (real line) and improved algorithm with multifractal (dotted line).

Another improvement is the better robustness on noise. We chart the average errors of Vn00 to Vn07 in Table 2, the curves are shown in Figure 9.

The higher the noise level becomes, the greater the accuracy improves. The improved method with multifractal is improved less sensitive to noise, and can be used to contain the deterioration caused by high noise.

Results of real image Rn and Rt have been compared to some manual segmenting results, and they match each other. The improved method can be well used for real applications.

4 CONCLUSIONS

An improvement from multifractal analysis has been done to the traditional tissue classifying algorithm using Fuzzy MRF Model. The original mathematical models and fuzzy features are reserved, when spatial resolution is increased, thus accuracy is improved. In numbers of tests on various sorts of data, the improved method shows its advantage on accuracy to the original method. Also an entire algorithm using the improved method is proposed and tested, doing well in real applications.

ACKNOWLEDGEMENTS

This work is supported by Project NSFC-60372023, the National Natural Science Foundation of China. We would like to give special thanks to Su Ruan, Jean-Marc Constans and Lab of GREYC-ENSICAEN CNRS UMR 6072 of France for providing us their research experience and experimental code.

REFERENCES

Rajapakse, J.C., Giedd, J.N., DeCarli, C., Snell, J.W., McLaughlin, A., Vauss, Y.C., Krain, A.L., Hamburger, S., Rapoport, J.L., 1996. A technique for single channel MR brain tissue segmentation: application to a pediatric sample. *Magn Res Imag*, 14, 1053~1065.

Taxt, T., Lundervold, A., 1994. Multispectral analysis of the brain using magnetic resonance imaging. *IEEE Trans Med Imag*, 13, 470~481.

Suzuki, H., Toriwaki, J., 1991. Automatic segmentation of head MRI images by knowledge guided thresholding. *Comput Med Imag Graph*, 15, 233~240.

Joliot, M., Mazoyer, B., 1993. Three-dimensional segmentation and interpolation of magnetic resonance brain images. *IEEE Trans Med Imag*, 12, 269~277.

Simmons, A., Arridge, S.R., Barker, G.J., Cluckie, A.J., Tofts, P.S., 1994. Improvement to the quality of MRI cluster analysis. *Magn Res Imag*, 12, 1191~1204.

Chang, M.M., Sezan, M.I., Tekalp, A.M., Berg, M.J., 1996. Bayesian segmentation of multislice brain magnetic resonance imaging using threedimensional Gibbsian Priors. *Optical Engineering*, 35, 3206~3221.

Ruan, S., Jaggi, C., Xue, J., Fadili, J., Bloyet, D., 2000. Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling. *IEEE Trans Med Imag*, 19, 1179~1187.

Ruan, S., Moretti, B., Fadili,J., and Bloyet, D., 2001. Segmentation of Magnetic Resonance Images using Fuzzy Markov Random Fields. *IEEE*, 3, 1051~1054.

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I., 1986. Fractal measures and their singularities, the characterization of strange sets. *Phys. Rev.*, 33, 1141–1151.

Sarkar, N., Katsuragawa, S., 1995. Multifractal and generalized dimension of gray-tone digital images. *Signal Processing*, 42, 181~190.

Liu, Y., Li, Y., 1997. New approaches of multifractal image analysis. *IEEE-ICICS' 1997*, 2, 970~974.

Ruan, S., Bloyet, D., 2000. MRF Models and Multifractal Analysis for MRI Segmentation. *IEEE-ICSP'2000*, 2, 1259~1262.

NONLINEAR MODELING OF CARDIOVASCULAR RESPONSE TO EXERCISE

Lu Wang[1], Steven W. Su[2,1], Gregory S. H. Chan[1], Branko G. Celler[1]*, Teddy M. Cheng[1] and Andrey V. Savkin[1]

[1] *Biomedical System Lab, School of Electrical Engineering & Telecommunications, Faculty of Engineering University of New South Wales, UNSW Sydney, N.S.W. 2052, Australia*

** b.celler@unsw.edu.au*

[2] *Key University Research Centre for Health Technologies, Faculty of Engineering, University of Technology Sydney,PO Box 123 Broadway NSW 2007 Australia*

Keywords: Cardiovascular system, Nonlinear modelling, Cardiovascular responses to Exercise, Machine learning.

Abstract: This study experimentally investigates the relationships between central cardiovascular variables and oxygen uptake based on nonlinear analysis and modeling. Ten healthy subjects were studied using cycle-ergometry exercise tests with constant workloads ranging from 25 Watt to 125 Watt. Breath by breath gas exchange, heart rate, cardiac output, stroke volume and blood pressure were measured at each stage. The modeling results proved that the nonlinear modeling method (Support Vector Regression) outperforms traditional regression method (reducing Estimation Error between 59% and 80%, reducing Testing Error between 53% and 72%) and is the ideal approach in the modeling of physiological data, especially with small training data set.

1 INTRODUCTION

The relationships between central cardiovascular variables and oxygen uptake during steady state of graded exercise have been widely examined by numerous investigators (Allor et al., 2000) (Astrand et al., 1964) (Fairbarn et al., 1994) (Freedman et al., 1955) (Kobayashi et al., 1978) (Reeves et al., 1961) (Richard et al., 2004) (Rowland et al., 1997) (Turley et al., 1997). Most of them investigated the relationship between cardiac output (CO) and oxygen uptake ($\dot{V}O_2$) using linear regression methods and found the slope between the two variables to be approximately 5 – 6 in normal and athletic subjects (Rowell et al., 1986). Beck et al (Beck et al., 2006) in contrast, investigated this relationship in healthy humans using polynomial regression. Turley (Turley et al., 1997) described both the relationship of stroke volume (SV) and the total peripheral resistance (TPR) to oxygen uptake during steady state of sub-maximal exercise using linear regression. However, from the point view of modeling, the regression methods used by the previous researchers have several limitations. First the empirical risk minimization (ERM) principle used by traditional regression models does not guarantee good generalization performance and may produce models that over-fit the data (Gunn, 1997). Secondly, most of the regression models developed from early research based on a small sample set with limited subjects during three or four exercise intensities. Traditional regression approachs are particularly not recommended for modeling small training sets. Determination of the size of the training set is a main issue to be solved in the modeling performance because the sufficiency and efficiency of the training set is one of the most important factors to be considered.

This study presents a novel machine learning approach, Support Vector Regression (SVR) (Drucker et al., 1997) to model the central cardiovascular response to exercise. SVR, developed by Vapnik and his co-workers in 1995, has been widely applied in forecasting and regression (Su et al., 2007) (Su et al., 2005) (Su et al., 2006) (Valerity et al., 2003). The following characteristics of SVR make it an ideal approach in modeling of cardiovascular system. Firstly, SVR avoids the over-fitting problem which exists in the traditional modeling approaches. Second, SVR condenses information in the training data and provide a sparse

representation by using a small number of data points (Girosi, 1998). Thirdly, SVR is insensitive to modeling assumption due to its being a non-parametric model structure. Finally, the SVR model is unique and globally optimal, unlike traditional training which can risk converging to local minima.

The rest of this paper is organized as follows: section 2 describes the experimental design for the data collection. Section 3 applies SVR for modeling the relationships between central cardiovascular variables and oxygen uptake. Finally, some conclusions are drawn in Section 4.

2 EXPERIMENTAL DESIGN

2.1 Subjects

We studied 12 normal male subjects. They are all active, but do not participate in formal training or organized sports. However, since two of them could not complete 6 minutes of higher level exercise, only the data recorded from 10 subjects (aged 25 ± 4yr, height 177 ± 5cm, body weight 73 ± 11kg) are used for this study. All the subjects knew the protocol and the potential risks, and had given their informed consent.

2.2 Experimental Procedure

All tests were conducted in the afternoon in an air-conditioned laboratory with temperature maintained between 23-24 ^{o}C. The subjects were studied during rest and a series of exercise in an upright position on an electronically braked cycle ergometer. Exercise was maintained at a constant workload for 6 minutes, followed by a period of rest. The initial exercise level was 25W and each successive stint of exercise was increased in 25W steps until a workload of 125W was reached. The rest periods were increased progressively from 10 to 30 minutes after each stint of exercise. Six minutes of exercise was long enough to approach a steady state since the values of oxygen uptake and the A-V oxygen difference had become stable by the 5th and 6th minutes even for near maximum exertion (Reeves et al., 1961).

2.3 Measurement and Data Processing

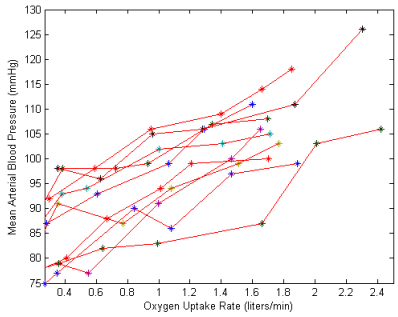
Heart rate was monitored beat by beat using a single lead ECG instrument, while ventilation and pulmonary exchange were measured on a breath by breath basis. Minute ventilation was measured during inspiration using a Turbine Flow Transducer model K520-C521 (Applied Electrochemistry, USA). Pulmonary gas exchange was measured using S-3A and CD-3A gas analyzers (Applied Electrochemistry, USA). Before each individual exercise test, the turbine flow meter was calibrated using a 3.0 liters calibration syringe. Before and after each test, the gas analyzers were calibrated using reference gases with known O_2 and CO_2 concentrations. The outputs of the ECG, the flow transducer and the gas analyzers were interfaced to a laptop through an A/D converter (NI DAQ 6062E) with a sampling rate of 500 Hz. Programs were developed in Labview 7.0 for breath by breath determination of pulmonary gas exchange variables but with particular reference to $\dot{V}O_2$ ($\dot{V}O_2$ STPD). Beat by beat stroke volume and cardiac outputs were measured noninvasively using the ultrasound based device (USCOM, Sydney, Australia) at the ascending aorta. This device has previously been reported to be both accurate and reproducible (Knobloch et al., 2005). In order to keep consistent measurements, all CO/SV measurements were conducted by the same person. An oscillometric blood pressure measurement device (CBM-700, Colin, France) was used to measure blood pressure.

The measurement of $\dot{V}O_2$ and HR were conducted during the whole exercise and recovery stage. The static values ($\dot{V}O_2$ and HR) were calculated for each workload from data collected in the last minute of the six minute exercise protocol. The measurements of SV, CO and BP (blood pressure) were similarly conducted during the last minute of the six minute exercise for each workload with the additional requirement that subjects keep their upper body as still as possible to minimize artifacts caused by the movement of the chest during exercise. We then, calculated their static values (CO, SV and BP) based on the measurement in the last minute for each workload.

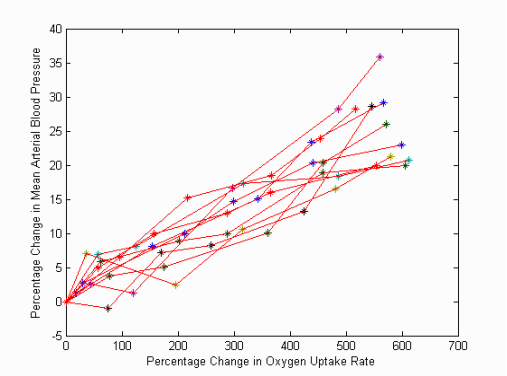
2.4 Results

We found that the percentage changes of cardiovascular variables relative to their rest values more uniform than when absolute values are used. This may be because using relative values diminish the variability between subjects. For example, Figure 1 (a) shows the relationship between the absolute value of mean arterial blood pressure response and the absolute value of oxygen uptake rate for all the ten subjects, while Figure 1 (b) is the percentage change in mean arterial blood pressure relative to its rest value with the percentage change in oxygen uptake rate to its rest value for the ten

subjects. Obviously, the response in Figure 1 (b) is more consistent and gives clearer trend than that in Figure 1 (a). It is thus reasonable to believe that modeling of cardiovascular responses using relative changes may give more robust results than modeling with the absolute values.



(a). Mean arterial blood pressure response to oxygen uptake rate (in absolute value).



(b). Mean arterial blood pressure response to oxygen uptake rate (in relative value).

Figure 1: Mean arterial blood pressure response to oxygen uptake rate for the ten subjects.

Based on the above finding, we model CO, SV and TPR to $\dot{V}O_2$ by modeling the percentage changes in CO, SV and TPR with respect to their corresponding rest values to percentage change in $\dot{V}O_2$ with respect to its rest value. We use CO%, SV%, TPR% and $\dot{V}O_2$% to represent their relative values (expressed as percentage), respectively.

3 APPLICATION OF SVR FOR MODELING

We selected radial basic function (RBF) kernels for this study, that is $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$ where σ is the kernel parameter, x_i is the ith input support value and x is the input value.

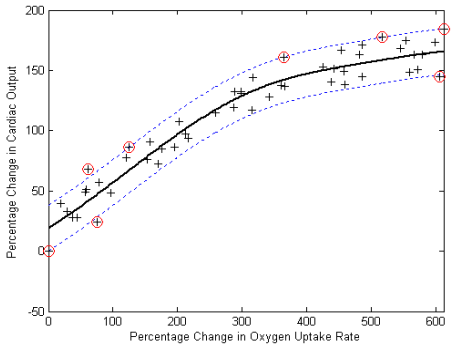
Detailed discussion about SVR, such as the selection of regularization constant C, radius ε of the tube and kernel function, can be found in (Gunn, 1997) (Vapnik, 1998).

In order to show the effectiveness of SVR, we applied both SVR and traditional linear regression (Least-Square linear regression (LS)) to investigate the relationships between percentage change of cardiovascular variables (CO%, SV% and TPR%) and $\dot{V}O_2$%.

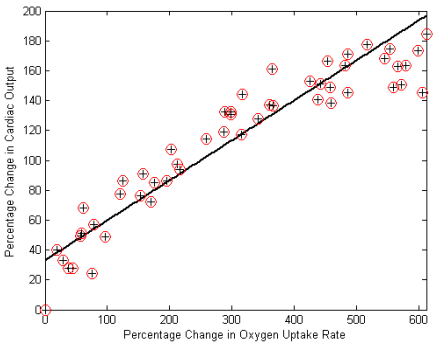
3.1 The Relationship between CO% and $\dot{V}O_2$%

3.1.1 Model Identification

A SVR model was developed to estimate CO% from $\dot{V}O_2$% (Table 1 and Figure 2). Although it is widely accepted that there is a linear relationship between cardiac output and oxygen consumption (Allor et al., 2000) (Astrand et al., 1964) (Freedman et al., 1955),



(a). Estimation of percentage change in CO from percentage change in $\dot{V}O_2$ using SVR.



(b). Estimation of percentage change in CO from percentage change in $\dot{V}O_2$ using LS.

Figure 2: Comparison of estimation results of CO% between using SVR and using LS.

their relationship can be better described by the nonlinear SVR model in terms of reducing the errors (MSE) from 418 to 171 (Table 2), an improvement of 59% comparing with that of LS method

The results in Table 1 also show the efficiency of SVR. Unlike traditional regression method where the solution of the model depends on the whole training data points, in SVR, the solution to the problem is only dependent on a subset of training data points which are referred to as support vectors. Using only support vectors, the same solution can be obtained as using all the training data points. SVR uses just 13% of the total points available to model their nonlinear behavior efficiently.

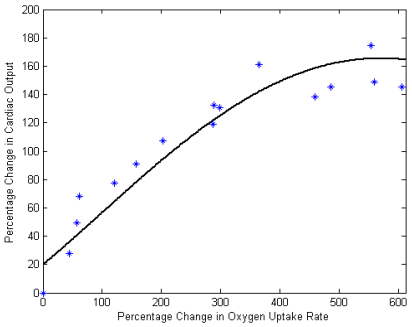
3.1.2 Model Validation

To further evaluate the feasibility of this proposed SVR model, the whole data set is divided into two parts: the first part (70% of the data) is used to design the model and the second part (30% of the data) is used to test its performance. Because we do not have large sample of data, we separated the data set into two parts randomly five times. Each time we use 70% of the data for training and the rest for testing. We established the SVR model with the three design parameters (kernel function, capacity (C) and the radius of insensitivity (ε)) based on the training set, and test its goodness on the testing set. In Figure 3, we present the results for one of the 5 tests. As shown in Table 3, the averaged results (MSE) for the 5 times testing for SVR is 245±15. However, the averaged error for traditional linear regression is as high as 521±19. It indicates that SVR can build more robust models to predict CO% from $\dot{V}O_2$% using only a small training set. It also demonstrates that SVR can overcome the over-fitting problem, even though SVR has more model parameters than the traditional linear regression method.

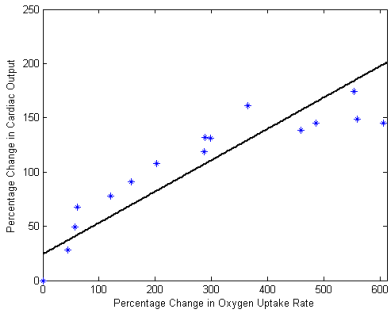
3.2 The Relationship between SV% and $\dot{V}O_2$%

Figure 4 shows the models for estimating SV%. The SVR model gives more precisely estimation than the LS does and decreases estimation errors (MSE) by 67% (Table 2).

The testing models are given in Figure 5 and the testing errors are in Table 3. As indicated, the SVR model decreases the testing error by 64%.

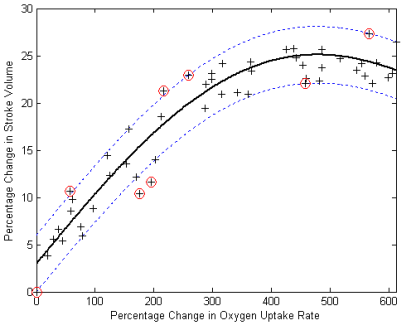


(a). Testing of SVR model.

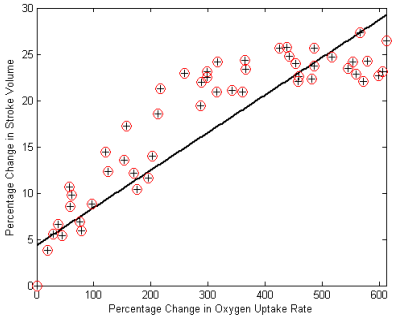


(b). Testing of LS model

Figure 3: Comparison of models of CO% against % change in oxygen uptake using SVR and using LS methods.

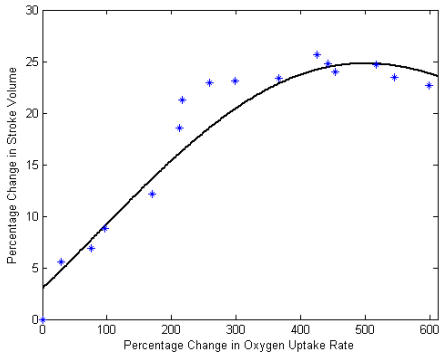


(a). Estimation of percentage change in SV from percentage change in $\dot{V}O_2$ using SVR.

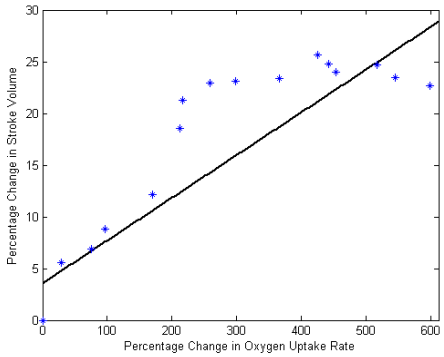


(b). Estimation of percentage change in SV with percentage change in $\dot{V}O_2$ using linear regression.

Figure 4: Comparison of estimation results for SV% between using SVR and using LS.



(a). Testing of SVR model.



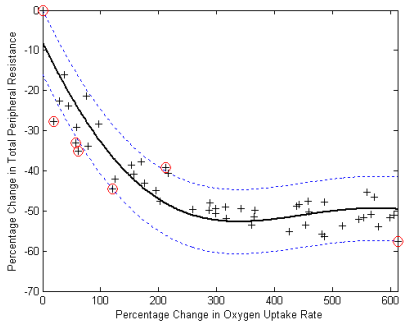
(b). Testing of LS model.

Figure 5: Comparison of the testing results for Stroke Volume using SVR and using traditional linear regression.

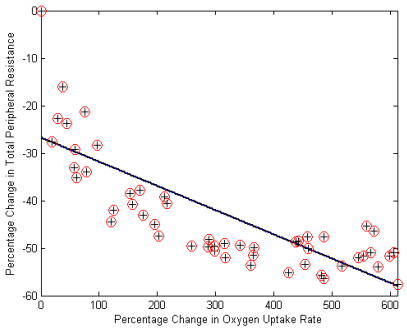
3.3 The Relationship between TPR% and $\dot{V}O_2$%

As shown in Figure 6, the SVR model describes a rapid fall in TPR% at low workloads which remains relatively constant even with increasing $\dot{V}O_2$%. SVR uses just 13% (Table 1) of the total points to get an efficient nonlinear model. Compared with linear regression, the SVR model decreases MSE from 151 to 30, an improvement of 80%.

The testing results for this SVR model and the equivalent LS model are given in Figure 7 and Table 3, respectively. Both of these (Figure 7 and Table 3) demonstrate that SVR outperforms the traditional linear regression method by reducing testing errors significantly, from 130 to 36.

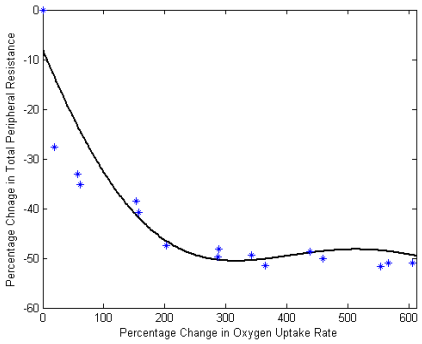


(a). Estimation of percentage change in TPR from percentage change in $\dot{V}O_2$ using SVR.

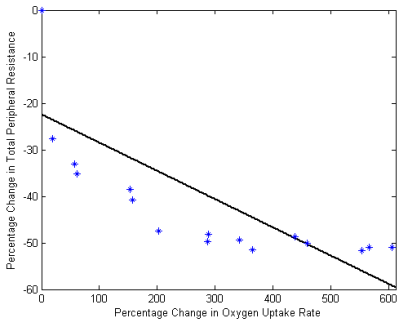


(b). Estimation of percentage change in TPR with percentage change in $\dot{V}O_2$ using linear regression.

Figure 6: Comparison of the estimation results of TPR% between using SVR and LS.



(a). Testing of SVR model.



(b). Testing of LS model.

Figure 7: Comparison of the test results of TPR% against % change on Oxygen uptake using SVR and using LS.

Table 1: Fitting data for the model of cardiovascular variables and oxygen uptake rate using SVR.

Relation	CO% Vs $\dot{V}O_2$%	SV% Vs $\dot{V}O_2$%	TPR% Vs $\dot{V}O_2$%
Kernel	RBF	RBF	RBF
Parameter	σ = 200	σ = 500	σ = 500
Regularization constant C	5000	5000	5000
ε-insensitivity	19	3	8
Support vectors number	8 (13.3%)	8 (13.3%)	8 (13.3%)
Estimation error	171	5	30

Table 2: Comparison of the estimation errors (MSE) between using SVR and using linear regression method

Relation	CO% Vs $\dot{V}O_2$%	SV% Vs $\dot{V}O_2$%	TPR% Vs $\dot{V}O_2$%
SVR	171	5	30
LS	418	15	151

Table3: Comparison of the model fitting errors (MSE) using SVR and linear regression methods (N=5).

Relation	CO% Vs $\dot{V}O_2$%	SV% Vs $\dot{V}O_2$%	TPR% Vs $\dot{V}O_2$%
SVR testing error	245 ± 15	8 ± 2	36 ± 5
LS Testing error	521 ± 19	22 ± 7	130 ± 12

4 CONCLUSIONS

This is the first time that SVR has been applied to experimentally investigate the steady state relationships between key central cardiovascular variables and oxygen consumption during incremental exercise. The impressive results obtained prove that SVR is an effective approach that can be recommended for the modeling of physiological data.

ACKNOWLEDGEMENTS

The authors acknowledge project DP0452186/2003 funded by ARC (Australian Research Council).

REFERENCES

Allor KM, Pivarnik JM, Sam LJ et al (2000) Treadmill economy in girls and women matched for height and weight. *J. Appl. Physiol* 89: 512-516.

Astrand PO, Cuddy TE, Saltin, B et al (1964) Cardiac output during submaximal and maximal work. *J. Appl. Physiol* 9: 268-274.

Beck KC, Randolph LN, Bailey KR et al (2006) Relationship between cardiac output and oxygen consumption during upright cycle exercise in healthy humans. *J. Appl. Physiol* 101: 1474 – 1480.

Drucker H, Burges C, Kaufman L et al (1997) Support vector regression machines. In: *Advances in Neural Information Procession Systems,* edited by M. Mozer, M. Jordan & T. Petsche. Cambridge, MA. p. 155-161.

Fairbarn MS, Blackie SP, McElvaney NG et al (1994) Prediction of heart rate and oxygen uptake during incremental and maximal exercise in healthy adults. *Chest* 105: 1365-1369.

Freedman ME, Snider GL, Brostoff P et al (1955) Effects of training on response of cardiac output to muscular exercise in athletes. *J. Appl. Physiol* 8:37-47.

Girosi F (1998) An equivalence between sparse approximation and support vector machines. *Neural Computation,* 20, 1455-1480.

Gunn SR (1997) Support Vector Macines for classificationa and regression, Technical Report, Department of Electronics and Computer Science, University of Southampton.

Knobloch K, Lichtenberg A (2005) Winterhalter M, Rossner D, Pichlmaier M,. and Philips R. Non-invasive cardiac output determination by two-dimensional independent Doppler during and after cardiac surgery. *Ann Thorac Surg* 80: 1479-1483.

Kobayashi Y, Andoh Y, Fujinami T et al (1978) Impedance cardiography for estimating cardiac output during submaximal and maximal work. *J. Appl. Physiol* 45: 459-462.

Reeves JT, Grover RF, Blount SG Jr et al (1961) Cardiac output response to standing and treadmill walking. *J. Appl. Physiol* 16: 283-288.

Reeves JT, Grover RF, Filley GF et al (1961) Circulatory changes in man during mild supine exercise. *J. Appl. Phsiol* 16: 279-282.

Richard R, Lonsdorfer-wolf E, Dufour S et al (2004) Cardiac output and oxygen release during vintensity exercise performed until exhaustion. *Eur J Appl Physiol* 93: 9-18.

Rowell LB. Circulatory adjustments to dynamic exercise (1986) In: Human circulation regulation during physical stress. New York: Oxford University Press, p. 213-256.

Rowland T, Popoeski , Ferrone L (1997) Cardiac response to maximal upright cycle exercise in healthy boys and men. *Med Sci Sport Exer* 29:1146-1151.

Su SW, Wang L, Celler BG (2006), Estimation of Oxygen Consumption for Moderate Exercises by Using a Hammerstein Model, In *Proceeding of the 28th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society, New York, USA, September, pp. 3427-3430.

Su SW, Wang L, Celler BG, Savkin AV, and Guo Y (2007). Identification and Control for Heart Rate Regulation during Treadmill Exercise, *IEEE Transactions on Biomedical Engineering* 54 (7): 1238-1246.

Su SW, Wang L, Celler B et al (2005) Estimation of walking energy expenditure by using support vector regression. In *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS),* pp. 3526–3529, Shanghai, China.

Turley KR, Wilmore JH (1997) Cardiovascular responses to treadmill and cycle ergometer exercise in children and adults. *J. Appl. Physiol* 83: 948-957.

Valerity VG, Supriya BG (2003) Volatility forecasting from multiscale and high-dimensional market data. *Neurocomputing*, 55, 285-305.

Vapnik V (1998) *Statistical learning theory*. New York: Wiley.

ANALYSIS OF HEART RATE AND BLOOD PRESSURE VARIABILITY IN PREGNANCY

New Method for the Prediction of Preeclampsia

H. Malberg
Karlsruhe Research Center, Institute for Applied Computer Science, Herrmann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen, Germany
malberg@iai.fzk.de

R. Bauernschmitt
Clinic for Cardiovascular Surgery, German Heart Center Munich, Germany

T. Walther
Department of Cardiology, Charité, Campus Benjamin Franklin (CBF), Berlin, Germany

A. Voss
Department of Medical Engineering, University of Applied Sciences, Jena, Germany

Renaldo Faber, Holger Stepan
Department of Obstetrics and Gynecology, University of Leipzig, Germany

N. Wessel
Department of Physics, University of Potsdam, Germany

Keywords: Heart Rate Variability, Blood Pressure Variability, baroreceptor reflex; screening, risk stratification.

Abstract: Pre-eclampsia is a serious disorder with high morbidity and mortality occurring during pregnancy; 3%–5% of all pregnant women are affected. Although most pre-eclamptic patients show pathological uterine perfusion in the second trimester, this parameter has a positive predictive accuracy of only 30%, which makes it unsuitable for early, reliable prediction. The study is based on the hypothesis that alterations in cardiovascular regulatory behavior can be used to predict PE. Ninety-six pregnant women in whom Doppler investigation detected perfusion disorders of the uterine arteries were included in the study. Twentyfour of these pregnant women developed PE after the 30th week of gestation. During pregnancy, additional several non-invasive continuous blood pressure recordings were made over 30 min under resting conditions by means of a finger cuff. In the period between the 18th and 26th weeks of pregnancy, three special variability and baroreflex parameters were able to predict PE several weeks before clinical manifestation. Discriminant function analysis of these parameters was able to predict PE with a sensitivity and specificity of 87.5% and a positive predictive value of 70%. The combined clinical assessment of uterine perfusion and cardiovascular variability demonstrates the best current prediction several weeks before clinical manifestation of PE.

1 INTRODUCTION

Pre-eclampsia (PE) is a serious pregnancy-specific disorder. It is characterized by sudden hypertension >140/90 mm Hg and a proteinuria (>300 mg in 24 hours). The manifestation of PE is the main cause of maternal and neonatal morbidity and mortality; it occurs in 3-5 % of all pregnancies.

Although the etiology and pathogenetic factors of the disease are largely unknown, early risk assessment by Doppler sonography has become an established procedure. However, the positive predictive accuracy (PPA) of Doppler sonography is limited to 30 %, as pregnant women with disturbed uterine perfusion may develop a PE, a pregnancy-induced hypertension (PIH), or a neonatal intrauterine growth retardation (IUGR) (Chien, 2000).

Earlier studies were unable to find either independent markers in the maternal plasma or physiological parameters easy to measure and, in this way, improve the screening efficacy of Doppler sonography (Benedetto, 1998).

Analyses of heart rate variability (HRV), systolic (SBPV) and diastolic blood pressure variability (DBPV) and baroreflex (BR) sensitivity (BRS) were able to demonstrate their high diagnostic and prognostic powers in various studies characterizing autonomous cardiovascular regulation in various diseases (La Rovere, 2001). Various studies demonstrated the suitability of these methods in hypertensive disorders of pregnancy, such as chronic hypertension (Walther, 2005), gestational hypertension (Hermida, 1998), and in PE (Faber, 2004). However, these diseases were clinically manifest already at the time of examination.

In contrast to those other studies, this study employs the approach of looking for characteristic alterations in cardiovascular regulation before the sudden rise of blood pressure. The study is based on the hypothesis that alterations in cardiovascular regulatory behavior can be used to predict PE. Conventional clinical prediction, i.e. Doppler sonography, has to be taken into account. Earlier findings have shown that the sole use of variability analysis in the 18th – 22nd weeks of gestation (WOG) was able to attain a PPA of 50 % (Walther, 2006). In addition, a combined study of variability and uterine perfusion achieved a PPA of 71.6 %, which may be considered the best finding for a non-invasive risk marker of PE at this point in time. In the study outlined below, findings are to be validated in an extended group of patients over a longer period of examination between the 18th and the 26th weeks of gestation.

2 PATIENTS

96 patients with abnormal uterine perfusion (AUP) were included in the study. All pregnant women underwent Doppler sonography in the 2nd trimester of pregnancy (median 22nd week of gestation, WOG, range 18 – 26 weeks) at the Department of Obstetrics and Gynecology of the University of Leipzig. 24 of these pregnant women developed PE after the 30th week of gestation. Approval by the local ethics committee and the informed consent of all subjects were obtained. All pregnancies were singleton. At the time of examination, the women were healthy, normotensive, without clinical signs of cervical incompetence, and on no medication.

Clinically, the development of pregnancy was subdivided in accordance with PE, pregnancy-induced hypertension (PIH), intrauterine growth retardation (IUGR), or pregnancy without any further complications. PE was classified in line with the guidelines of the International Society for the Study of Hypertension in Pregnancy. PIH was described by the rise of several blood pressure levels to more than 140 mm Hg in the systole and more than 90 mm Hg in the diastole within four hours. Significant proteinuria is characterized by an excretion of more than 300 mg of total protein in 24 hours. Where these data were not available, proteinuria was detected by dipstick on two consecutive occasions within four hours. Intrauterine growth retardation was defined by the birth weight being below the 10th percentile of a reference group.

3 METHODS

Doppler examination of the uterine arteries was carried out with a LOGIQ 9 ultrasound machine (GE, Solingen, Germany) with a 5 MHz convex transducer by the same sonographer. Uterine perfusion was classified as pathological when there was bilateral notching or a mean pulsatility index (PI) of both arteries above 1.45. Immediately after the Doppler examination, continuous blood pressure monitoring was conducted non-invasively via finger cuff (100 Hz, Portapres device mod. 2, BMI-TNO, Amsterdam, The Netherlands). The measurements were performed under standardized resting conditions between 8 a.m. and 12 noon. The continuous blood pressure curves were used to extract the time series of beat-to-beat intervals, systolic and diastolic blood pressures.

3.1 Preprocessing

The main objective of the analysis of heart rate and blood pressure is to investigate the cardiovascular system during normal sinus rhythm. Therefore, it is necessary to exclude not only artifacts (e.g. double recognition, i.e. R-peak and T-wave recognized as two beats) but also beats not coming from the sinus node of the heart, so called ventricular premature complexes (VPC). VPCs are not directly controlled by the autonomous nervous system. Practically, this exclusion means filtering of the time series. The original time series are denoted RR-series (derived from the RR-intervals) and the filtered time series, NN-series (normal-to-normal beat interval, NNI). VPCs in the tachogram are usually characterized by

regular ventricular premature beat and supraventricular premature beat complexes). The 20%-filter (Kleiger, 1987) considers these facts; if the current value of the tachogram differs from its predecessor by more than 20%, the current value and its successor are marked not normal. VPCs with less than 20% difference are not removed from the series and may falsify almost all HRV or BPV parameters. The RR-intervals recognized as not normal are treated in different ways: either they are simply removed from the series or interpolated linearly or spline interpolated (Lippmann, 1994). Interpolating linearly may lead to false decreased variability's, interpolating with splines often fails in time series with many VPCs. In several clinical studies, an adaptive filtering algorithm introduced in (Wessel, 2000) has been proven to exclude premature beats and artifacts. The main advantage of this procedure is the spontaneous adaptation to variability changes in the series, which enables a more reliable removal of artifacts and VPCs. This new filtering algorithm consists of three sub-procedures: (i) the removal of obvious recognition errors, (ii) the adaptive percent filter, and (iii) the adaptive controlling filter. A MATLAB implementation of the preprocessing algorithm is available from <tocsy.agnld.uni-potsdam.de>.

3.2 Heart rate and Blood Pressure Variability

Standard methods of HRV analysis include time and frequency domain parameters; these are linear methods. Time domain parameters are based on simple statistical methods derived from the RR-intervals as well as the differences between them. The mean heart rate is the simplest parameter, but the standard deviation over the whole time series (sdNN) is the most prominent HRV measure for estimating overall HRV. A list of these parameters is given in Table 1. These parameters can be calculated for short (5 minutes) and long (24 hours) term epochs, representing short-term and long-term variability, respectively, or for averaged short-term epochs (e.g. a mean of 288 five-minute intervals a day). The overall HRV estimate, sdNN, and other time domain parameters can be used to predict mortality in the recovery period after myocardial infarction. In one of the first risk studies using these parameters, (Kleiger, 1987) showed that an sdNN<50ms was associated with a 5.3-fold increased mortality when compared to patients with preserved HRV (sdNN>100ms).

Time domain geometric methods (see Table 1) are methods by which the NNIs are converted into special geometric forms quantifying their distribution. Special forms are used to make the approach more insensitive to artifacts and ectopic beats. A disadvantage of these methods is that they require a considerable number of RR-intervals; they are thus not applicable to very short-term time series. A triangular index, HRVi, showing reduced HRV has been associated with both arrhythmic and non-arrhythmic death (Task Force, 1996).

Table 1: Description of time and frequency domain parameters, adopted standards (Task Force, 1996) and additional measures developed by the authors (•). NNI stands for the filtered beat-to-beat intervals (NN-intervals).

Variable	Units	Definition
Time domain statistical methods		
meanNN	ms/mm Hg	Mean NNI and mean BP, respectively
sdNN	ms/mm Hg	Standard deviation of all NNI and BP values, respectively
rmssd	ms/mm Hg	Root mean square of successive NNI/BP differences
•pNNX	%	Percentage of beat-to-beat differences greater than X ms/mm Hg (e.g. X = 3/6/9 ms/mm Hg)
•pNNlX	%	Percentage of beat-to-beat differences lower than X ms/mm Hg (e.g. X = 3/6/9/12 ms/mm Hg)
•Shannon	None	Shannon entropy of the histogram (density distribution of the NNIs/ BP values)
•RenyiX	None	Renyi entropy of the order X of the histogram (e.g. X = 2/4/0.25)
Time domain geometric methods		
HRVi	none	HRV triangular index (see (Task Force, 1996) for details)
TINN	ms	Baseline width of the minimum square difference triangle
Frequency domain methods		
P	$ms^2/mm\ Hg^2$	Total power from 0 − 0.4Hz
VLF	$ms^2/mm\ Hg^2$	Very low frequency band, 0.0033 − 0.04Hz
LF	$ms^2/mm\ Hg^2$	Low-frequency band, 0.04 − 0.15 Hz
HF	$ms^2/mm\ Hg^2$	High-frequency band, 0.15 − 0.4 Hz
LF/HF	None	Quotient of LF and HF
LFn	None	Normalized low-frequency band (LF/(LF+HF))
cross 1/f		Intercept of a log-log-power spectrum
slope 1/f		Slope of a log-log-power spectrum

We introduced a more robust method to quantify the distribution (Voss, 1996) based on information theory, in particular the Shannon and the Renyi entropies of the histogram. We demonstrated the usefulness for risk stratification in a blinded study two years later (Voss, 1998). Frequency domain HRV parameters allow periodic dynamics in the heart rate time series to be analyzed (Akselrod, 1981). There are mainly two different techniques for spectral analysis: methods based on Fast Fourier Transform (FFT) and parametric autoregressive model estimates of wavelet approaches. The results obtained from using different spectral methods should be comparable though (apart from differences in time and frequency resolution). The Task Force on HRV (Task Force, 1996) recommended that power spectral analysis of 5-minute ECG recordings be used to assess autonomic physiology and pharmacology.

Very low, low and high frequencies (see Table 1) can be estimated from such 5-minute ECG recordings. In this study, all frequency domain parameters were calculated from the complete 30-minute recording.The high frequency power reflects the modulation of vagal activity by respiration whereas the low-frequency power represents vagal and sympathetic activities via the baroreflex loop. The low-to-high frequency ratio is used as an index of sympathovagal balance (Malliani, 1991). The suitability of frequency domain parameters for risk stratification of post-infarction patients was proven by Bigger et al. (Bigger, 1992) - a reduction in ultra low and very low frequency power is associated with pathologies.For blood pressure series, all HRV parameters described can be calculated accordingly, only some statistical parameters need to be adapted (e.g. pNN50 makes no sense for BPV - the standard deviation for BP series varies between 5 and 10 mmHg).

3.3 Baroreflex Sensitivity

Analysis of spontaneous baroreflex sensitivity (BRS) is very important for cardiac risk stratification of various cardiovascular diseases (La Rovere, 2001). BRS slope is defined as the instinctive change of NNI related to increasing or decreasing levels of systolic blood pressure and is expressed in [ms/mmHg]. There is evidence showing that a decreased BRS may carry an adverse prognosis in cardiac patients (La Rovere, 1998).

For several years, BRS was determined pharmacologically (phenylephrine, nitro-prusside) (Vanoli, 1994) or mechanically (Cohen, 1981) until, in the 1980s, innovative methods of estimating BRS were developed based on spontaneous heart rate and blood pressure fluctuations (Di Rienzo, 1985).

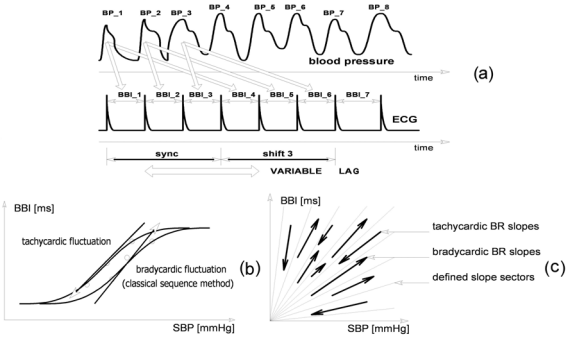


Figure 1: The Dual Sequence Method of estimating spontaneous BRS analyses simultaneous (sync) and delayed responses (shift 3, variable delay) of heart rate to blood pressure increases (a) as well as bradycardic and tachycardic blood pressure fluctuations (classical sequence method (b). Moreover, also the slope sector distribution is quantified (c). These slope sectors also can be defined as overlapping regions.

These methods evaluate arterial baroreflex function in the absence of external stimulations of the cardiovascular system, therefore defined as “spontaneous”. These spontaneous techniques nowadays are the state of the art in research, though not in clinical practice. We introduced the Dual Sequence Method (DSM) (Malberg, 2002) for advanced spontaneous baroreflex sensitivity estimates. This method considers not only bradycardic (blood pressure increase causes RR-interval increase) and tachycardic (blood pressure decrease causes RR-interval decrease) blood pressure fluctuations as introduced in the sequence method (Di Rienzo, 1985) (see Figure 1 (a)), but also defines slope sectors quantifying the BRS slope distribution (see Figure 1 (b)). Earlier studies showed that the heart rate does not simultaneously respond to the blood pressure fluctuation (Manicia, 1985). Therefore, DSM quantifies synchronous as well as delayed heart rate response to the same BP fluctuation (see Figure 1 (c)).

In summary, these are the parameter blocks and ranges calculated by DSM:

(i) the total number of slopes in different sectors within the time series,

(ii) the percentage of slopes relative to the total number of slopes in different sectors,(iii) the numbers of bradycardic and tachycardic slopes,

(iv) the shift operation from the first to the third heart beat triples, a variable lag, and

(v) the average slope of all fluctuations and its standard deviation.

The average BRS slope is defined as the NNI difference related to SBP changes, and is estimated by linear regression.

The parameters, 'P_brady', and, 'P_tachy', characterize the incidence of increasing and decreasing SBP triples with regard to the total number of SBP values. Consequently, these parameters estimate the basic cause of BRS activity. A reduced number of ramps in SBP unavoidably leads to a reduced number of HR responses. The parameters are defined as

*P_brady = (No. of increasing SBP triples/ total No. of SBP triples) * 100%*

*P_tachy = (No. of decreasing SBP triples/ total No. of SBP triples) * 100%*

The percentage of adequate HR responses (BRS events) relative to the numbers of SBP ramps is described by the 'Activation' parameter. It is defined as

*Activation = (No. of BRS events/ No. of SBP ramps) * 100%.*

In contrast to classical BRS methods, DSM defines slope sectors allowing to quantify the BRS slope distribution. Sectors with a range of 2 ms/mm Hg have been proven to act as a suitable partition in patient studies. Then, the percentages of BRS events in different slope sectors relative to the total number of BRS can be estimated. Moreover, the total number of BRS events is normalized to the mean heart rate. For detailed definitions of the DSM parameters, reference is made to the original contribution (Malberg, 2002). These parameters are calculated for bradycardic as well as tachycardic fluctuations, both synchronous or delayed, to analyze a possibly delayed response of the heart rate to the same blood pressure oscillation. This DSM method is used to quantify sequences of length three; longer sequences turned out not to be useful for spontaneous BRS estimates because of their low occurrence.

3.4 Statistics

In this study, the Kruskal-Wallis test was used to determine intergroup differences in clinical parameters. The Mann-Whitney U test was employed to analyze the differences in variability parameters among pregnant women with uterine perfusion disorders developing PE (number = 24) compared to those not developing PE (NoPE, number = 72). The level of significance of the intergroup differences was defined as $p < 0.05$. Due to the explorative character of the study we did not apply the multiple test correction. Stepwise discriminant analysis was employed to determine the best combinations of parameters for separating individual groups.

4 RESULTS

In this study, no pregnant woman with normal uterine perfusion developed hypertensive pregnancy-related disorders. In the period between the 18th and the 26th weeks of gestation, in the abnormal uterine perfusion group, the following differences were found by variability analysis (Tables 2 – 3). In HRV analysis, both the mean and the standard deviation were unchanged, while some frequency domain parameters showed significant differences. Interestingly, all significant parameters point to very low frequencies below 0.04 Hz.

As in the HRV analysis, the mean values and the standard deviation were unchanged also in SBPV and DBPV. On the other hand, especially in DBP, time domain and frequency domain parameters as well as non-linear parameters showed significant differences. The most prominent difference was found to be the 'high frequency' in diastolic blood pressure.

Table 2: HRV analysis in the 18th – 26th weeks of gestation in pregnant women with abnormal Doppler findings developing either PE or NoPE after the 30th week of gestation.

	NoPE	PE	P value
meanNN	759.8±104.4	755.3±113.4	n.s.
sdNN	44.7±16.2	49.0±18.1	n.s.
VLF	10.18±11.45	13.46±11.46	0.013
VLF/P	0.35±0.11	0.44±0.10	0.005
ULF/P	0.21±0.15	0.14±0.08	0.029
cross 1/f	1.79±1.62	3.09±1.48	<0.001
slope 1/f	-0.77±0.56	-0.38±0.51	0.002

Table 3: Analysis of systolic and diastolic blood pressure variability in the 18th – 26th weeks of gestation in pregnant women with abnormal Doppler findings developing either PE or NoPE after the 30th week of gestation.

SBPV	NoPE	PE	P value
meanNN	122.4±16.1	128.6±13.2	n.s.
sdNN	7.81± 2.03	8.36±1.86	n.s.
Rmssd	2.66±0.56	3.02±0.81	n.s.
pNN2	0.27±0.11	0.34±0.13	n.s.
LF	0.14±0.09	0.16±0.07	n.s.
HF	0.03±0.02	0.05±0.03	0.021
WPSUM02	0.46±0.16	0.43±0.11	n.s.
PLVAR2	0.031±0.045	0.014±.0171	n.s.
DBPV	NoPE	PE	P value
meanNN	68.0±11.2	72.9±9.0	n.s.
sdNN	4.15±1.07	4.74±1.43	n.s.
Rmssd	1.89±0.33	2.18±0.59	0.033
pNN2	0.13±0.06	0.20±0.11	0.012
LF	0.05±0.03	0.07±0.04	0.011
HF	0.01±0.01	0.02±0.01	0.002
WPSUM02	0.47±0.14	0.41±0.13	0.049
PLVAR2	0.110±0.083	0.080±0.114	0.004

Also BR regulation as characterized by DSM showed differences in pregnant women developing PE compared to women without PE (Table 4). Analysis reveals that the number of rises in SBP potentially initiating BR increases significantly in PE.

Table 4: Baroreflex analysis by the dual sequence technique in the 18th – 26th weeks of gestation in pregnant women with abnormal Doppler findings developing either PE or NoPE after the 30th week of gestation.

	Bradycardic BR Regulation		
Parameters	**NoPE**	**PE**	**P**
No. of SBP ramps	458.6±94.1	528.1±128.5	0.005
No. of SBP ramps [%]	18.5±2.6	21.1±4.0	<0.001
BR time delay [beats]	1.7±0.4	1.8±0.3	n.s.
No. of slopes between 4-6 ms/mm Hg	35.4±13.6	46.6±22.2	0.019
No. of slopes between 3-5 ms/mm Hg	36.5±15.2	49.1±20.4	0.004
No. of slopes between 5-7 ms/mm Hg	32.2±12.3	41.1±19.2	n.s.
Total No. of BR slopes	173.9±50.0	216.7±77.9	0.009
average BR slope [ms/mm Hg]	13.3±5.6	13.2±5.6	n.s.
BR Activation [%]	38.3±8.7	41.3±10.9	n.s.
	Tachycardic BR Regulation		
Parameters	**NoPE**	**PE**	**P**
No. of SBP ramps	464.9±106.3	527.0±97.3	0.008
No. of SBP ramps [%]	18.8±3.4	21.2±2.7	0.005
BR time delay [beats]	1.5±0.3	1.7±0.5	0.029
No. of slopes between 4-6 ms/mm Hg	41.9±19.3	60.6±20.6	<0.001
No. of slopes between 3-5 ms/mm Hg	47.0±22.7	65.3±24.8	0.003
No. of slopes between 5-7 ms/mm Hg	38.4±17.4	50.2±17.6	0.004
Total No. of BR slopes	200.0±59.1	242.0±64.6	0.005
average BR slope [ms/mm Hg]	12.9±5.6	12.2±6.1	n.s.
BR Activation [%]	43.5±9.4	46.5±11.5	n.s.

Nevertheless, BR per se does not change in its response to activation and mean rise. Also, the total number of BR fluctuations and the number in the low regulation segment (between 3 and 7 ms/mm Hg) is elevated in PE. In the tachycardic range, BR regulation also starts later in the PE group than in the NoPE group.

The application of stepwise discriminant analysis selected the following three parameters as the best parameters predicting PE: HRV: VLF/P, DBP: HF, BR: number of tachycardic slopes between 4-6 ms/mm Hg. A sensitivity and a specificity of 87.5 %, a positive predictive accuracy of 70.0 % were found with a negative predicted accuracy of 95 %. Interestingly, these are the same parameters which had been found in our previous study (Walther, 2006).

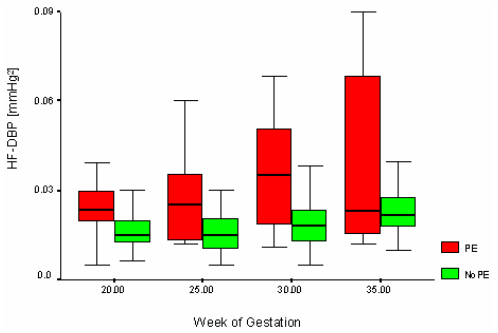


Figure 2: Example of the change in variability measure during pregnancy in the case of PE genesis. The increase in the high-frequency component of diastolic blood pressure is shown.

Figure 2 shows an example of the change over time of variability measure in the course of pregnancy in patients developing PE as against pregnant women without this development.

5 DISCUSSION

Early prediction of PE is one of the most important challenge in obstetrics. Establishing a simple test manageable under clinical conditions is a major challenge. Doppler sonography or combined with humoral or endothelial parameters either attained low sensitivity / sufficient sensitivity or a low positive predictive value and are very costly or even invasive.

Although BRV, BPV, and BRS have been established in cardiology for risk stratification, their use for early detection of hypertensive pregnancy disorders is still very rare. As various cardiovascular dis-

eases can be predicted by a gradual change in cardiovascular regulation, the approach used in this study also was to cover the genesis of PE. This study is aimed at investigating alterations in HRV, BPV and BRS to predict the sudden steep increase of blood pressure which is caused by PE.

With an incidence of 3 – 5 % of all pregnancies in the Western population, pathological uterine perfusion in the second trimester is known as an indirect sign of inadequate trophoblasts. The positive predicted value of this study, however, is only around 30 %. In an earlier study of the variability analysis of PE, the authors were able to show that the use of only the variability parameter was able to raise to 50 % the positive predictive value, which does represent the highest possible PPA, but is not yet sufficient for clinical routine screening. Except for that, the three variability parameters described above attained the highest PPA of any one method of examination which, in addition, is independent of humoral factors and other single clinical parameters.

This study has shown that the combination with Doppler sonography of uterine arteries confirms the highest possible PPA as compared to all published non-invasive trials (Walther, 2006). The PPA of approx. 70% is indicative of the clinical relevance of Doppler examination combined with variability analysis. However, due to the exploratory design of this study, these parameters need to be validated prospectively - especially in connection with humoral factors. Anyway, the same three parameters, which had been found in our previous study (Walther, 2006), were selected in the discriminant analysis. These parameters obtained nearly the same classification results – which is already a first validation.

On the basis of the variability measures determined, the following interpretation could be possible in the course of pregnancy for the genesis of PE in a cardiovascular model (see Figure 3).

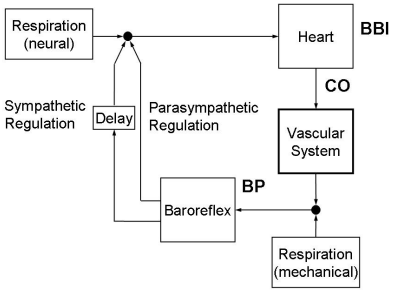


Figure 3: Connections between the heart rate, NNI, the cardiac output, CO, and the blood pressure, BP, in autonomous regulation of the cardiovascular system, modified from [Saul, 1991)

In Figure 3 a simplified model of the cardiovascular circulatory system is presented. Mediated by the electromechanical coupling, the NNIs initiate a cardiac output (CO) of the heart. The resulting blood pressure in the periphery is influenced by this CO and by the vascular system. In the feedback the NNIs are influenced by the BRS via sympathetic and parasympathetic activation. Additionally, the total system is superimposed by mechanical and neuronal influences of the respiration.

Our analyses point to a more vascular disorder as the cause of the cardiovascular alteration. Obviously, the heart plays only a secondary role. Apparently, there is an early pathological modification in vessel behavior measurable already in the Doppler sonogram. However, this parameter obtains a low PPA only. This incipient endothelial dysfunction, which is still very weak in the 18th – 26th weeks of gestation, has no influence on the mean values of diastole, systole, and heart rate. On the other hand, especially the variability of blood pressure seems to change as a consequence of continuing pathological arterial stiffness in so far as the minor fluctuations in blood pressure become more pronounced as a result of the decreasing windkessel function of vessels. This can be represented by the changes in blood pressure variability (see parameters: SBPV: ‘HF’, DBP: ‘Rmssd’, ‘pNN2’, ‘LF’, ‘HF’, ‘WPSUM02’, ‘PLVAR2’, BRS: ‘No. of SBP ramps’, ‘No. of SBP ramps [%]’). The baroreflex reacts more strongly to this change, i.e. it reacts more frequently to these slight, but more frequent blood pressure stimuli (see BR parameters: ‘No. of slopes between 4-6 ms/mm Hg’, ‘No. of slopes between 3-5 ms/mm Hg’, ‘No. of slopes between 5-7 ms/mm Hg’, ‘Total No. of BR slopes’). The baroreflex function (i.e. the intensity of response of the BR) seems to be completely unchanged (see BR parameters: ‘average BR slope’, ‘BR activation’). The changes in BPV and BR are continued in HRV either as a consequence of the counter regulation of the heart rate responding to blood pressure fluctuations, or due to other regulatory influences modulated by respiratory sinus arrhythmia (see HRV parameters: ‘VLF’, ‘VLF/P’, ‘ULF/P’, ‘cross 1/f’, ‘slope 1/f’). Thus, for example, the increase in diastolic high frequency, which is modulated by respiratory sinus arrhythmia, may reflect early pathological arterial stiffness. This leads to the undamped, respiration-induced pulse-wave oscillations detected by our method. This is congruent with the hypothesis that patients later developing PE are characterized by early pathological modifications in vessel behavior. The physiological conclusion could be drawn that variability analysis meas-

ures the consequences to blood pressure, to the interaction between blood pressure and heart rate, and to the heart rate of incipient endothelial dysfunction, which is not thought to be sufficient to predict PE solely on the basis of the Doppler sonogram.

In a previous methodological study the applicability of different BR methods was tested (Vanoli, 1994, Laude, 2004). All described methods estimate only the average BRS slope. The DSM however, is able to obtain additional insights into the cardiovascular regulation. In this study, 'average slope' is not altered, however more sophisticated DSM parameters found high significant differences. So we conclude, that the parameter 'average slope' is not sufficient for PE prediction. The best discrimination had been obtained by the combination of non-linear BR parameters and linear HRV und BPV parameters.

In summary, it can be said that examination of uterine perfusion combined with the characterization of cardiovascular regulation in the second trimester has achieved the most accurate prediction of PE several weeks before its clinical manifestation so far. In this application, the biosignal analysis emphasizes its importance as a non-invasive, cheap and universal diagnostic approach. This opens up potential therapeutic strategies for suppressing pathophysiological symptoms of the disease to further decrease maternal and neonatal morbidity and mortality rates.

REFERENCES

Chien, P.F., Arnott, N., Gordon, A., Owen, P., Khan, K.S. 2000, How useful is uterine artery Doppler flow velocimetry in the prediction of pre-eclampsia, intrauterine growth retardation and perinatal death? An overview. *Brit J Obstet Gynaecol* 107:196-208.

Benedetto, C., Valensise, H., Marozio, L., Giarola, M., Massobrio, M., Romanini, C. 1998, A two-stage screening test for pregnancy-induced hypertension and preeclampsia. *Obstet Gynecol* 92:1005-1011.

La Rovere, M.T., Pinna, G.D., Hohnloser, S.H., Marcus, F.I., Mortasa, A., Nohora, R., Bigger, T., Camm, A.J., Schwartz, P.J. 2001, Baroreflex sensitivity and heart rate variability in the identification of patients at risk for life-threatening arrhythmias. *Circulation* 103: 2072–2077.

Walther, T., Wessel, N., Baumert, M., Stepan, H., Voss, A., Faber, R. 2005, Longitudinal analysis of heart rate variability in chronic hypertensive pregnancy. *Hypertension Res* 28:113-118.

Hermida, R.C., Ayala, D.E., Mojon, A., Fernandez, J.R., Silva, I., Ucieda, R., Iglesias, M. 1998, Blood pressure excess for the early identification of gestational hypertension and preeclampsia. *Hypertension* 31:83-89.

Faber, R., Baumert, M., Stepan, H., Wessel, N., Voss, A., Walther, T. 2004, Baroreflex sensitivity, heart rate and blood pressure variability in hypertensive pregnancy disorders. *J Hum Hypertens* 18:707-712.

Walther, T., Wessel, N., Malberg, H., Voss, A., Stepan, H., Faber, R. 2006, A combined technique for predicting pre-eclampsia: concurrent measurement of uterine perfusion and analysis of heart rate and blood pressure variability. *Journal of Hypertension* 24:747-750.

Kleiger, R.E., Miller, J.P., Bigger, T., Moss, A.J. 1987, Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol* 59:256–262.

Lippman, N., Stein, K.M., Lerman, B.B. 1994, Comparison of methods for removal of ectopy in measurement of heart rate variability. *Am J Physiol* 267:H411-418.

Wessel, N., Voss, A., Malberg, H., Ziemann, Ch., Voss, H.U., Schirdewan, A., Meyerfeld, U., Kurths, J. 2000 Nonlinear analysis of complex phenomena in cardiological data. *Herzschr Elektrophys* 11: 159-173

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. 1996, Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation* 93:1043-1065.

Voss, A., Kurths, J., Kleiner, H.J., Witt, A., Wessel, N., Saparin, P., Osterziel, K.J., Schurath, R., Dietz, R. 1996, The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. *Cardiovasc Res* 31:419–433.

Voss, A., Hnatkova, K., Wessel, N., Kurths, J., Sander, A., Schirdewan, A., Camm, A.J., Malik, M. 1998, Multiparametric analysis of heart rate variability used for risk stratification among survivors of acute myocardial infarction. *Pacing Clin Electrophysiol* 21(1 Pt 2):186–192.

Akselrod, S., Gordon, D., Ubel, F.A., Shannon, D.C., Barger, A.C., Cohen, R.J. 1981, Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science* 213:220–222.

Malliani, A., Pagani, M., Lombardi, F., Cerutti, S. Cardiovascular neural regulation explored in the frequency domain. *Circulation*1991;84:482–92.

Bigger Jr, J.T., Fleiss, J.L., Steinman, R.C., Rolnitzky, L.M., Kleiger, R.E., Rottman, J.N. 1992, Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation* 85:164–71.

La Rovere, M.T., Bigger, J.T., Marcus, F.I., Mortara, A., Schwartz, P.J. 1998, Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction. atrami (autonomic tone and reflexes after myocardial infarction) investigators. *Lancet* 351(9101):478–484.

Vanoli, E., Adamson, P.B. 1994, Baroreflex sensitivity: methods, mechanisms, and prognostic value. *Pacing Clin Electrophysiol* 17(3 Pt 2):434–445.

Cohen, I.M., O'Connor, D.T., Preston, R.A., Stone, R.A. 1981, Long-term clonidine effects on autonomic function in essential hypertensive man. *Eur J Clin Pharmacol* 19(1):25–32.

Di Rienzo, M., Bertinieri, G., Mancia, G., Pedotti, A. 1985, A new method for evaluating the baroreflex role by a joint pattern analysis of pulse interval and systolic blood pressure series. *Med Biol Eng Comput* 23(suppl I):313-314.

Malberg, H., Wessel, N., Schirdewan, A., Hasart, A., Osterziel, K.J., Griessbach, G., Voss, A. 2002, Advanced analysis of the spontaneous baroreflex sensitivity using the Dual Sequence Method. *Clinical Science* 102: 465-473.

Saul, J.P., Berger, R.D., Albrecht, P., Stein, S.P., Chen, M.H., Cohen, R.J. 1991, Transfer function analysis of the circulation: unique insights into cardiovascular regulation. *Am J Physiol* 261: H1231-H1245.

Laude, D., Elghizi, J.L., Girard, A., Bellard, E., Bouhaddi, M., Castiglioni, P., Cerutti, C., Cividjan, A., Di Rienzo, M., Fortrat, J.O., Janssen, B., Karemaker, J.M., Lefthériotis, G., Parati, G., Persson, P.B., Porta, A., Quintin, L., Regnerd, J., Rüdiger, H., Stauss, H.M. 2004, Comparison of various techniques to estimate spontaneous baroreflex sensitivity (the EuroBaVar study). *Am J Physiol Regul Integr Comp Physiol*, 286: R226-231

STATISTICAL SIGNIFICANCE IN OMIC DATA ANALYSES

Alternative/Complementary Method for Efficient Automatic Identification of Statistically Significant Tests in High Throughput Biological Studies

Christine Nardini, Luca Benini
DEIS, University of Bologna, Viale Risorgimento 2, Bologna, Italy
cnardini@deis.unibo.it, lbenini@deis.unibo.it

Michael D. Kuo
UCSD Medical Center HillCrest, 200 West Arbor Drive, San diego, CA, USA
mkuo@ucsd.edu

Keywords: Statistical testing, statistical significance, multiple hypothesis testing, false discovery rate, statistical resampling methods, statistical meta-analysis, omic data.

Abstract: The post-Genomic Era is characterized by the proliferation of high-throughput platforms that allow the parallel study of a complete body of molecules in one single run of experiments (*omic* approach). Analysis and integration of *omic* data represent one of the most challenging frontiers for all the disciplines related to *Systems Biology*. From the computational perspective this requires, among others, the massive use of automated approaches in several steps of the complex analysis pipeline, often consisting of cascades of statistical tests. In this frame, the identification of statistical significance has been one of the early challenges in the handling of *omic* data and remains a critical step due to the multiple hypotheses testing issue, given the large number of hypotheses examined at one time. Two main approaches are currently used: p-values based on random permutation approaches and the False Discovery Rate. Both give meaningful and important results, however they suffer respectively from being computationally heavy -due to the large number of data that has to be generated-, or extremely flexible with respect to the definition of the significance threshold, leading to difficulties in standardization. We present here a complementary/alternative approach to these current ones and discuss performances and limitations.

1 INTRODUCTION

In recent times high-throughput devices for genome-wide analyses have greatly increased in size, scope and type. In the post-Genomic Era, several solutions have been devised to extend the successful approach adopted for gene expression analyses with microarray technology to other bodies of data such as proteomes, DNA copy number, single nucleotide polymorphisms, promoter sites and many more (Nardini et al., 2006). These data supports, and notably their integration, represent the future of molecular biology; for this reason the elucidation and definition of tools and methods suited to handle the data produced by these high-throughput devices is of great importance.

Early methods for such analyses were mainly dealing with gene expression data, their goal being to extract items that appear to have coherent trends among themselves (in this context commonly called *unsupervised* methods) or with respect to external features, such as clinical markers (*supervised* methods). Both types of approaches have been used for example for the classification of subtypes of poorly understood diseases with unpredictable outcomes (Ramaswamy et al., 2003; Lapointe et al., 2004). Currently, other approaches, that take advantage of larger and diverse sources of information are being devised to address questions of varying complexity in different areas of research rooted in molecular biology. These methods cover a broad variety of applications, from the study of complex hereditary diseases (Rossi et al., 2006) to the identification of radiological traits' *surrogate markers* (the molecular origin of a clinical trait) for enabling non-invasive personalized medicine (Segal et al., 2007). Overall, besides the variety and complexity of the analyses and methods adopted, some invariants can be identified. The most common atomic step is the identification on the large scale of similarities or associations among molecular behaviors. Such association measures consist for example of scores that evaluate similarities across several samples of genes' expression profiles, or genetic coherence in genes copy number or deletion, and more. Coherence among expression profiles and other association

measures can be assessed by means of statistical techniques, namely, by computing a measure of trend similarity (test score, θ) and evaluating the likelihood of this measure to occur by chance (α-level or p-value). The test score is then assumed to be either a measure of actual similarity or only a random effect, based on the value of the associated p-value. The p-value represents the probability of being wrong when assuming that the score represents an actual similarity. This error (type I error) can happen for non-extreme values of the test θ that are difficult to classify as *good* or *bad* and results in erroneously refuting the null hypothesis ($H_0 : \theta = 0$) which assumes that there is no relationship, when actual facts show that the items are tightly related. The scientific community typically assumes to be meaningful (i.e. *statistically significant*) test scores that are coupled to p-values lower or equal to one of the following nominal p-values: $0.05, 0.01, 0.001$. These values represent the probability of committing typeI errors. Given these definitions, the highly dimensional nature of genome-wide data has posed problems and challenges to conventional biostatistical approaches. Indeed, when performing in parallel such a large number of tests, typeI errors inherently rise in number, since over a large number of items, the possibility of faults increases. For this reason, p-values need to be readjusted in a more conservative way, accounting for the so called *multiple hypothesis testing* issue. The most classical technique to account for this problem is the Bonferroni correction (R.R.Sokal and F.J.Rohlf, 2003) that simply multiplies the actual p-value of every single test by the total number of tests observed. However, this approach is not considered viable in *omic* studies, as in fact it often leads to the rejection of too many tests, since none of the corrected p-value are smaller than any of the nominal p-values. An alternative and less conservative approach to this problem is the generation of a random distribution, based on random resampling or on the generation of scores obtained from the randomization of the data. Such approaches allow to build a distribution that represents the population's behavior, and can thus be used to test the hypothesis of interest. When operating with *omic* data, another statistic, the False Discovery Rate (FDR) has been introduced (Benjamini and Hochberg, 1995; Storey and Tibshirani, 2003; Tusher et al., 2001). Like the p-value, the FDR measures the false positives, however while the p-value controls the number of false positive over the number of truly null tests, the FDR controls the number of false positive over the fraction of significant tests. The utility of this statistic is undeniable, however, its interpretation is far less standardized than the better known p-value, and thus, very often, the value of acceptance of a test based on FDR is much more flexible and dependent on the investigator experience. Globally, these characteristics make the results assessed by FDR highly dependent on the rejection level the investigator chooses. This makes it difficult to automate with high parallelism the identification of statistically significant hypotheses. This problem can becomes relevant due to the increasingly common necessity to merge different sources of information to assess the validity of a given biological hypothesis. Examples of such circumstances arise whenever, for example, the analysis aims at refining, by means of cascades of statistical tests, a set of genes candidate to explain a biological assumption. The hypothesis in fact is refined collecting information across various databases or other forms of *a priori* knowledge, that progressively filter out the spurious data -only as an example see various tools presented in (Tiffin et al., 2006; Rossi et al., 2006). To be efficient, the analysis requires the result of each filtering step to be automatically sent to the following one. Thus the possibility to assess significance by mean of universally accepted values of significance becomes relevant. This latter observation was one of the stimuli motivating the search for an alternative/integrative approach to the multiple hypotheses problem encountered when dealing with genomic datasets. We also wanted this method to be reasonably efficient to be computed. We thus approached the problem based on techniques that allow the intrinsic correction of p-values in case of multiple tests (*meta analyses* approaches) used for the combination of various statistical tests. Among them, we turned our attention to the category of the *omnibus tests* (L.B.Hedges and I.Olkin, 1985). These approaches are non-parametric, meaning that they do not depend on the distribution of the underlying data, as long as the test statistic is continuous. In fact, p-values derived from such tests have a uniform distribution under the null hypothesis, regardless of the test statistic or the distribution they have been derived from. However, omnibus tests suffer from a strong limitation: they can be used to assess whether there is a superior outcome in *any* of the studies performed. This means that the combined significance is not a measure of the average significance of the studies performed. An omnibus test therefore cannot be used *as is*, to assess the global statistical validity of the number of tests considered simultaneously. Thus, we manipulated this approach to make it applicable to the definition of a significance threshold.

The main advantage of our solution is twofold. On one side the p-values can be computed in very reasonable times and can thus help managing the computational issues related to permutations techniques; on

the other side they represent p-values for which nominal threshold of significance (e.g. $0.05, 0.01, 0.001$) can be applied, and can overcome the threshold selection issue faced when using FDR approaches. Additionally, this method appears to perform slightly better than other methods in avoiding the selection of false positives. However, this is coupled to a partially diminished ability in identifying correctly true positives in complex patterns of association. These consideration support the findings of several authors that strongly suggest to validate the results obtained from *omic* studies through the use of different techniques and threshold of significance, given the highly noisy nature of the data (Pan et al., 2005).

2 RELATED WORK

Two main methodologies are currently being used to approach the multiple hypothesis testing issue. The first is based on the principles that define the resampling statistical approaches (R.R.Sokal and F.J.Rohlf, 2003). In particular we adopted the permutation method that requires the construction of a null distribution to which to compare the actual data. This distribution must be built from the generation of a large number of random data. When the distribution is built using the randomized data generated by *all* the tests, the corresponding p-value is corrected for these same multiple hypotheses. This represents a structurally simple, robust, but computationally intensive approach, given the large numbers involved in the analysis of *omic* data. The computational efficiency issue can become extremely relevant, since most of the interpreted languages commonly used for their large libraries of bioinformatics related functions (notably R and the Bioconductor Project (Gentleman et al., 2005), and Matlab), cannot reasonably handle such approaches. Even with the recent improvements for (implicit) parallelization of the computation, time lags for the evaluation of the results remain large. Moreover, for large datasets, compiled languages such as C also require intensive and long lasting computational efforts, unless specific architectures are adopted to enhance efficiency. The second approach consists of novel methods purposely introduced to handle *omic* data that defines the concept of False Discovery Rate. This statistic comes in a number of flavors, and relies on complex statistical assumption. A full description is beyond the scope of this paper, here we briefly describe three of the most used approaches: (i) the pioneering work of Benjamini (Benjamini and Hochberg, 1995); (ii) the definition of the q-value (Storey and Tibshirani, 2003); (iii) the FDR adopted in the tool Significance Analysis of Microarray -SAM, (Tusher et al., 2001)- a widespread software used for the analysis of microarray data.

Benjamini FDR: This approach controls the FDR by modifying the p-values obtained on a single test, rescaling it in the following way: $FDR_{BEN} = \frac{Kp_i}{i\sum_{i=1}^{K} i^{-1}}$, where p_i represents the i-th of the K single p-values.

q-value: The q-value is the minimum false discovery rate. This measure can be approximated by the ratio of the number of false positives over the number of significant tests, the implementation of the q-value provides several options to evaluate this estimate and to compare it to the corresponding p-values. $q \approx min(\#\text{false positives}/\#\text{significant tests})$.

SAM FDR: SAM is a tool that allows the extraction of significant genes that help differentiate 2 or more sample classes by means of various scores suited to answer different questions (i.e. depending on the number of sample classes observed and on the meaning of the scores defining the classes, such as survival times, experimental points in time course experiments etc.). Statistical validation of the score value produced by SAM is performed by the generation of a distribution of random score values. These scores are evaluated by means of random permutations of the class labels. These new values, along with the ones from the original classification are used to evaluate the FDR as the average of falsely significant items: $FDR_{SAM} = \frac{\frac{\#signif.\ permuted\ scores}{\#permutations}}{\#signif.\ actual\ scores}$ i.e. the number of items with permuted test scores called significant divided by the number of permutations over the number of items called significant in actual data.

The q-value approach is one of the most widespread, both because of its quality and because of the various and user-friendly implementations the authors have made available. For this reason we choose this method for comparison to ours. In general, FDR scores represent an extremely valuable information while dealing with *omic* data, however, the main issue to the fully automated use of these techniques lies in the flexible acceptance of the threshold values for significance. In other words the investigator can set his threshold for the acceptance of the False Discovery Rate, but no universally accepted thresholds have been recognized. This issue has been pointed out for example in (Cheng et al., 2004). In this work the authors designed three other statistical scores to help in the choice of the threshold for significance. Among these scores, two are designed to assess general significance threshold criteria for large-scale multiple tests and one is based on existing biological knowledge. Our method does not represent a novel way to

evaluate FDR, but it defines a p-value, for this reason universally accepted thresholds for significance can be adopted.

More recently and independently from our approach (Yang and Yang, 2006) have designed a method based on omnibus tests to improve the identification of the FDR. Again, one of our goals is to provide an efficient way to evaluate a p-value that takes into account the multiple hypotheses tested, in order to be able to adopt the thresholds of significance accepted by the scientific community (0.05,0.01,0.001), easier to automate in long pipelines of tests. In this paper we show that the p-value obtained with manipulation of the inverse χ^2 method (one of the omnibus tests) can also be used directly as a measure of significance for the identification of statistically significantly tests.

3 METHOD

We chose as the base for our approach the inverse χ^2 method (L.B.Hedges and I.Olkin, 1985), an *omnibus* statistical test used to ascertain if at least one among several tests is significant, by evaluation of the following statistics: $S(k) = -2\sum_{i=1}^{k} ln(p_i)$ and $s(k) = \chi^2(S,2k)$ where $k = 1...K$ are the tests performed and p_i the p-value of the i-th test. S has a $\chi^2_{(s,2k)}$ distribution, where s is the p-value of the χ^2 distribution with $2k$ degrees of freedom, and represents the significance of the combined tests, meaning that it can assess if *any* of the tests can be considered significant, accounting for the total number of K tests performed. Thus, in the following, s will indicate the p-value we can use for assessing the statistical significance of the tests taking into account the multiple hypothesis issue, while p will indicate the significance of the single test. The score θ is the value resulting from the statistical test. Making use of the χ^2 inverse method means testing the null hypothesis $H_0 : H_{0,1} = ... = H_{0,K} = 0$. Values of $s > 0.05$ indicate that H_0 cannot be rejected and thus that it holds for all the subhypotheses $H_{0,i} = 0, i \in [1,K]$. Conversely, more than one combination of rejection and non rejection of single hypotheses $H_{0,i}$ is possible to justify the rejection of the global null hypothesis H_0. For example all but one of the subhypotheses could be null, or only one could be null etc. Evaluating s on all the tests performed would be of no interest in terms of defining a global threshold for significance. In fact, while a non significant value of s would indicate that none of the items has a score value that allows the rejection of the null hypothesis, a low value of s (< 0.05) would only mean that at least one item's score is relevant to the rejection of the null hypothesis, with no indication on which one(s) are the relevant items. To overcome this limitation we ranked the tests scores θ in ascending order (assuming that significant values of the test are represented by high values of the score), and ordered the p-values consistently. We then evaluated s for sets of p-values of increasing size, starting from a set made of only the p-value corresponding to the worse test score, then adding at each iteration of this algorithm another p-value coupled to the immediately higher or equal (better) score (θ), and closing the last iteration with all the p-values. By induction (Equation 1) we can show that whenever the value of s drops below any of the standard values of significance $(0.05, 0.01, 0.001)$ the score corresponding to the last p-value added is the threshold for significance, since it represents the specific test that accounts for the impossibility to reject the global null hypothesis H_0. By construction, at each iteration, the p-value added is always smaller, and correspondingly, due to the logarithm properties, S shows a fast growth ($S(k) = -2\sum_{i=1}^{k} ln(p_i)$). At the same time the parameter of the χ^2 function k, grows linearly ($2 \cdot k$). Because of the shape of the χ^2 function and because of the logarithm properties, if there are *enough* small p-values, S becomes quickly and abruptly very large, and moves to behaviors typical of the ones on the right hand side of Figure 1(c), $\chi^2_k(S) \rightarrow_{k\rightarrow\inf, S\rightarrow\inf} 0$. This gives s its typical shape (shown in Figure 1(b)), with a very abrupt drop from values very close to 1 to values very close to 0.

$$
\begin{array}{lll}
For & i = 1 & s(i) > 0.05 \Rightarrow \\
 & H_0 \; not \; rej. & H_{0,1} \; not \; rej. \\
Let & i = n & s(i) > 0.05 \Rightarrow \\
 & H_0 \; not \; rej. & H_{0,i} \; not \; rej., \forall i \in [1,n] \\
Then & i = n+1 & s(i) > 0.05 \Rightarrow \\
 & H_0 \; not \; rej. & H_{0,i} \; not \; rej., \forall i \in [1,n+1] \\
 & & s(i) \leq 0.05 \Rightarrow \\
 & H_0 \; rej. & H_{0,i} \; not \; rej., \forall i \in [1,n], \\
 & & H_{0,n+1} \; rej.
\end{array}
\tag{1}
$$

Figure 1 shows an example of the trends of the variables involved in the evaluation of global significance: the statistics S and s that define the global significance, the test score θ and the corresponding single p-value that are the basic units of the analysis. The statistic S represents the argument of the χ^2 function and is associated to a given degree of freedom (k). For any given degree of freedom it is possible to identify the minimum value (here called $S_{id\alpha}$) for which the inverse χ^2 function returns the suited probability α. Since S_{id} is the minimum value, the p-value that represents the threshold for significance is associated to $k_{sign\alpha}$ and can be conveniently visualized as the point

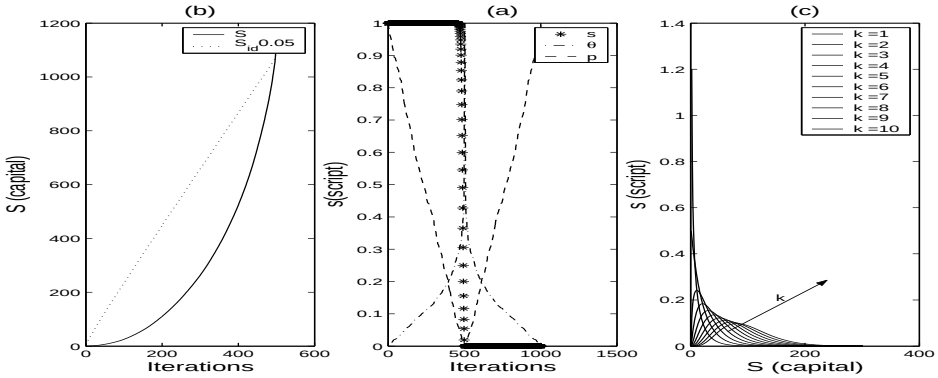


Figure 1: Graphical representation of the different scores involved in the analysis. Figure(a) deals with the statistic S and S_{id}. Figure(b) plots the corrected p-value s, the absolute value of the correlation score θ and the single p-value p. Figure(c) shows the χ^2 probability density function..

in which $k_{sign\alpha} = k|S_{id\alpha}(k) = S(k)$. Equivalently for s the threshold for significance at a given nominal level α can be defined as $k_{sign} = min_{k\in[1,K]}|s(k) \leq \alpha$. In our experiments θ is the Spearman correlation score (R.R.Sokal and F.J.Rohlf, 2003). Before processing the test values we separated positive from negative scores, and then performed the previously described operations on the absolute values. This sign segregation of the data has a two-fold objective. On one side this fulfills the requirement for the applicability of the test since one tailed p-values are required. On the other side it satisfies the biological necessity to discern between significantly over and under expressed genes, based on positive and negative values of the test scores. As far as the permutation approach is involved we generated 1000 random permutations of each trait values as it was done in other applications with this same goal (Liang et al., 2005). We then re-evaluate the θ scores for all 1000 randomized instance of each trait, these constitute the null distribution. For the FDR approach, we used the q-value R package with default settings. For the identification of significant items, we adopted as threshold the same values we used for the p-value. The method was implemented in Matlab, scripts for the method are available upon request.

3.1 Data

To test our method, we simulated the typical set up of a common genomic experiment. Namely, we generated a random expression matrix $1000x100$ (i.e. 1000 genes and 100 samples) and we defined 5 external traits for which we search the *surrogate markers*. In other words, these external traits mimic any clinical trait or molecular marker. The goal of the experiment is to identify the genes associated to the external traits, to define the traits' surrogate markers. This approach is then used to investigate the molecular etiology of commonly used clinical markers. Several examples of such approaches can be found in literature, only as a sample see (Lapointe et al., 2004; Liang et al., 2005). At first, we tested the method's ability to recognize surrogate markers of variable size. The surrogate markers were obtained either by simple copy of expression profiles (in varying number of copies, namely $0,1,5$), or by sum of varying numbers of profiles (namely $5,30$). The first group of external traits (#1,#2,#3) provides both the negative control (0 copies, obtained by elimination of a randomly chosen expression profile, and exported as external trait) and helps measuring the comparative ability of the 3 different approaches (FDR, permutations and our method) in extracting small cluster of correlated profiles ($1,5$ copies). The second set of traits (#4,#5) tests the approach with more challenging data (sums of $5,30$ copies). To each expression value we added varying levels of gaussian noise ($0\%, 50\%, 100\%$) proportional to the expression value, to better mimic real data (Bansal et al., 2007). To avoid specific case results, we replicated our approach 3 times per each noise level and averaged the results of specificity, sensitivity, positive and negative predictive value. We observed the approach for the 3 levels of significance 0.05,0.01, 0.001. Finally, we tested our method to assess its reliability with variable numbers of genes.

3.2 Multiclass Statistical Scores

To compare our results we evaluated the specificity, sensitivity, negative and positive predictive value of the 3 methods: permutations, FDR and ours. These statistics are used in combination to quantify different aspects of the accuracy of a binary test, evaluating different proportions of correctly and incorrectly classified items, when compared to a known classification, considered the gold standard. In this context the *test* is

the ensemble of all the operations performed to classify each items; *positive* and *negatives* label the items according to the two classes $c = N, P = 0, 1$ they belong to; *true* (T) and *false* (F) represent the ability of the test to classify coherently or not a given item in the test classification with respect to the gold standard classification. Thus, for example, in classical definitions TN (true negative) labels items belonging to class 0 (N) correctly classified by the test, and FP (false positive) labels items incorrectly classified as 1 (P) by the test. Given these definitions, positive and negative predictive value (PPV, NPV), sensitivity (Se) and specificity (Sp) are usually formalized with the relationships in the first part of Equations 2.

Table 1: Classical definition and generalization to 3 classes for *true, false, negatives, positives*.

(a) Classical Definition

Test \ Gold Standard	T	F	
P	TP	FP	$\rightarrow P_t$
N	FN	TN	$\rightarrow N_t$
	$\downarrow$	$\downarrow$	
	P_{gs}	N_{gs}	

(b) 3-Classes Definition

Test \ Gold Standard	2	1	0	
2	T_2	x_{12}	x_{13}	$\rightarrow C_{2,t}$
1	x_{21}	T_1	x_{23}	$\rightarrow C_{1,t}$
0	x_{31}	x_{32}	T_0	$\rightarrow C_{0,t}$
	$\downarrow$	$\downarrow$	$\downarrow$	
	$C_{2,gs}$	$C_{1,gs}$	$C_{0,gs}$	

When the test classifies $n > 2$ categories, these definitions become more complex to apply. However, it still remains important to be able to characterize the performances of the test in terms of its ability to distinguish between items that belong and do not belong to any category (in our case between genes that constitute and do not constitute any molecular surrogate). To reach this goal and preserve the meaning of the 4 scores (PPV, NPV, Se, Sp) some caution must be used. In fact the meaning of *positive* and *negative* is not relevant anymore, since there are now *positives*. Then, while the definition of *true* remains straightforward, as it indicates coherence between the classification of the test and the gold standard, the definition of *false* can be cumbersome, since there are $n-1$ ways to misclassify an item. Additionally, the possibly intuitive definition of *false positives* (or *negatives* as items that are non-zero in the test (or in the gold standard) classification leads to ambiguity, since items happen to be contemporary false positives *and* false negatives. To avoid confusion and ambiguities the actual values of all false can be identified by rewriting the problem in terms of a system of equation based on the relationships indicated in Table 1. Here P_t, N_t represent the total number of positive and negative items that can be found in the test (t) categorization, and P_{gs}, Ngs in the gold standard (gs) classification. The definitions can be generalized to $n > 2$ classes changing the term negative and positive with the indices of the corresponding classes $c = 0, 1, ..., n$, and having C_c that designs the total number of positives for each given class. The system of equations obtained from the relationships in the rows and columns of Table 1 contains $2 \cdot n$ equations (i.e. $TP + FP = P_t$) and $2 \cdot n$ unknown (x_{ij}), thus it is completely specified. It is worth noticing, that with these general definitions, in case of 2-classes test, Se and Sp appear to be dual scores. Thus, when generalizing to n-classes it is possible to define the predictive ability of the test for each given class $c \in 0, 1, .., n$ as $PV_c = T_c/C_t$ and the Sensitivity/Specificity (now called Sep) for the same class c as $Sep_c = T_c/C_{gs}$. To clarify the situation it is extremely useful to rewrite the definitions as they are written on the left hand side of Equation 2, namely:

$$\begin{array}{rcccl} PPV & = & TP/TP+FP) & = & TP/P_t \\ PPN & = & TN/(TN+FN) & = & TN/N_t \\ Se & = & TP/(TP+FN) & = & TP/P_{gs} \\ Sp & = & TN/(TN+FP) & = & TN/N_{gs} \end{array} \quad (2)$$

For n classes this gives:

$$\begin{array}{rcl} PPV & = & \sum_c T_c / \sum_c C_{c,t}, c = 1, .., n \\ PPN & = & T_0/N_t = T_0/C_{0,t} \\ Se & = & \sum_c T_c / \sum_c C_{c,gs}, c = 1, .., n \\ Sp & = & T_0/N_{gs} = T_0/C_{0,gs} \end{array} \quad (3)$$

4 RESULTS AND DISCUSSION

All the results obtained with our method were obtained in much more efficient times compared to the permutation method, since the computational complexity of our algorithm is $O(g \cdot t)$ while the bootstrapping one is $O(g \cdot t \cdot p)$, with g indicating the number of genes, t the number of external traits, and p the number of permutations. The comparison with FDR in these terms is not relevant, since this method is computationally efficient. We performed 3 main experiments: the first for comparison among the 3 methods across all the types of traits (global comparison, Table 2); then more specifically, trait by trait (Table 3); finally we explored the stability of the method across varying numbers of tests performed.

As far as the first comparison is involved, all methods performed with varying good degrees of specificity ($Sp > 0.95$), but none had satisfactory sensitivity ($Se < 0.5$ to $Se << 0.5$) except the permutation method for only the threshold 0.05, $Se_{perm,\alpha=0.05} = .67$. In particular, our method has intermediate sensitivity (better than FDR) and specificity (better than

Table 2: Statistics of the performances of the 3 methods compared: our method, permuted p-values and FDR. The comparison is done on expression matrices 1000x100 and 5 traits as they are described in Section 3.1. Results are averaged over 3 instances of the random data generated with the same specifics. Standard deviations of these averages are below 10^{-2}. The first column indicates the noise level (n), the second the threshold of significance chosen (α) and then all the scores for the 3 methods. Because of space constraints only values for noise 0.5 are shown.

		Our Method		Permutations		FDR - q-value	
n	α	Se	Sp	Se	Sp	Se	Sp
0.5	.05	.1905	.9998	.6746	.9512	.1667	.9948
	.01	.1667	.9999	.4603	.9898	.1667	.9948
	.001	.1667	1.000	.3175	.9981	.1667	.9948

Table 3: Class by class comparison of the algorithms performances. Our method performs better in terms of avoiding false positive ans worse with false negatives. Data are shown as averages across the random replicates and across the 3 different levels of significance, for 3 different levels of noise (n). Figures in italic were inferred from NANs.

		PV (*classes*)						Sep (*classes*)					
n	Method	0	1	2	3	4	5	0	1	2	3	4	5
0	Ours	.9998	1.000	*1.000*	1.000	.3111	.0556	.9936	1.000	*1.000*	1.000	*0.000*	*0.000*
	Perm.	.9797	1.000	*1.000*	1.000	.9556	.2852	.9956	.2510	0.000	.3846	.4325	.3494
0.5	Ours	.9999	1.000	*1.000*	1.000	.0444	.0037	.9931	1.000	*1.000*	1.000	*0.000*	*0.000*
	Perm.	.9797	1.000	*1.000*	1.000	.9556	.2852	.9956	.2510	0.000	.3846	.4325	.3494
1	Ours	.9999	.3333	*1.000*	.7333	.0000	.0037	.9925	*0.000*	*1.000*	.9506	*0.000*	*0.000*
	Perm.	.9797	1.000	*1.000*	1.000	.9556	.2852	.9956	.2510	0.000	.3846	.4325	.3494

permutations). Since the FDR method at the chosen thresholds for significance appears to behave in extreme ways, i.e. with better specificity and worse sensitivity with respect to both methods, we focused our attention to a more refined comparison between the bootstrapping method and ours, and did not pursue the goal, out of our scope here, to evaluate results with other thresholds for significance.

Namely, we performed the second experiment, on a trait by trait basis, with two goals: to investigate the reasons of the improved performances of our method in terms of specificity; to assess the reasons for the poor global performances in terms of sensitivity. For this we evaluated PV and Sep for each one of the 6 classes ($c = 0, 1, .., n$). In general our method seems to have more problems with false negatives, while the bootstrapping method collects a much larger number of false positives (Table 3). These characteristics depend on the intrinsic properties of s as they have been described in Section 3. The abrupt drop in value of s is responsible for an almost binary behavior of this score. This leaves very little *gray* areas for spurious classification, thus ambiguous θ values are quickly coupled to high s values and discarded from the significant tests set. Overall, trait #5 defines a too complex pattern (sum of 30 profiles), and none of the method can treat it correctly, conversely, trait #4 (sum of 5 profiles) can be superiorly handled by the permutation method and trait #1, #2 and #3 ($1, 0, 5$ correlated profiles) are better recognized with our method. It is difficult to speculate on whether surrogate markers of type #3 are more or less common than the ones

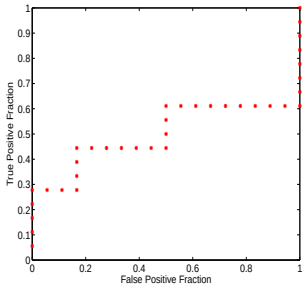


Figure 2: ROC curve for PV, AUC ≈ 0.6.

of type #4 in actual biology, we can state however that our method is able to identify the surrogate markers of trait #3 with profiles that have as little correlation as 0.33 (100% noise addedd). To summarize these results we evaluated ROC curves to assess if any of the methods was strikingly outperforming the other (ROC curves in this case are not used to evaluate the relationship between sensitivity and specificity, but to compare two populations of data, that happen to be PV and Sep scores). We compared: (i) PV and Sep for each method, (ii) Sep only, (iii) PV only. Namely, sensitivity and specificity combined, as well as sensitivity alone lead to $AUC \approx 0.5$, while the specificity test leads to $AUC \approx 0.6$, slightly better, but not statistically significant (Figure 2, $AUC = 0.5$ indicates tests with comparable performances).

Finally, we tested our method for the same hypotheses for varying numbers of genes, from 100 to 2000 (steps of 100 genes). Across 20 samples we obtained median values that reproduce the findings of

the two previous experiments (global and trait by trait performances) with very small variances across the 20 samples ($\approx 10^{-2}$ for sensitivity and $\approx 10^{-3}$ for specificity). Thus, the method appears to be stable with respect to the number of items tested.

5 CONCLUSIONS

We presented a method for the identification of *p*-values in *omic* studies. This approach is based on a meta-analysis and has two main advantages. On one side it is computationally efficient, and can thus be used in interpreted languages such as R and Matlab that offer rich libraries of functions for *omic* analyses. On the other side it is based on the identification of a *p*-value rather than FDR, and can thus take advantage of nominal threshold for significance, allowing for an easier automation of filtering steps in analyses based on statistical tests. Conversely to the permutation technique, that remains a computationally intensive but very robust reference method, our approach, globally, appears to be more specific but less sensitive. This improved specificity can be extremely advantageous in the practice of Systems Biology, since novel compact functional subunits can emerge or remain uncovered and require longer and costly experimental investigations to be extracted, depending on the noise they appear to be identified with. Application to real data needs to be provided and this represents our current research activity. For these reasons we believe the definition of alternative and complementary method is appropriate.

ACKNOWLEDGEMENTS

The authors would like to thank Diego di Bernardo and Mukesh Bansal for constructive discussion.

REFERENCES

Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and di Bernardo, D. (2007). How to infer gene networks from expression profiles. *Mol Syst Biol*, 3.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J.R. Stat. Soc. B*, 57:289–300.

Cheng, C., Pounds, S., Boyett, J., Pei, D., Kuo, M., and Roussel, M. F. (2004). Statistical significance threshold criteria for analysis of microarray gene expression data. *Stat Appl Genet Mol Biol*, 3:Article36.

Gentleman, R., Carey, V., Huber, W., Irizarry, R., and Dudoit, S. (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.

Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci.*, 101(3):811–816.

L.B.Hedges and I.Olkin ((1985)). *Statistical Methods in Meta-Analysis*. Academic Press, New York.

Liang, Y., Diehn, M., Watson, N., Bollen, A. W., Aldape, K. D., Nicholas, M. K., Lamborn, K. R., Berger, M. S., Botstein, D., Brown, P. O., and Israel, M. A. (2005). Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc. Natl. Acad. Sci.*, 102(16):5814–5819.

Nardini, C., Benini, L., and Micheli, G. D. (2006). Circuits and systems for high-throughput biology. *Circuits and Systems Magazine, IEEE*, 6(3):10–20.

Pan, K.-H., Lih, C.-J., and Cohen, S. N. (2005). Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc. Natl. Acad. Sci.*, 102(25):8961–8965.

Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003). A molecular signature of metastasis in primary solid tumors. *Nat. Genet.*, 33(1):49–54.

Rossi, S., Masotti, D., Nardini, C., Bonora, E., Romeo, G., Macii, E., Benini, L., and Volinia, S. (2006). TOM: a web-based integrated approach for efficient identification of candidate disease genes. *Nucleic Acids Res.*, 34(doi:10.1093/nar/gkl340):W285–W292.

R.R.Sokal and F.J.Rohlf (2003). *Biometry*. Freeman, New York.

Segal, E., Sirlin, C. B., Ooi, C., Adler, A. S., Gollub, J., Chen, X., Chan, B. K., Matcuk, G. R., Barry, C. T., Chang, H. Y., and Kuo, M. D. (2007). Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nature Biotechnology*, 25(6):675–680.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *PNAS*, 10(16):9440–9445.

Tiffin, N., Adie, E., Turner, F., Brunner, H., van Driel nd M. Oti, M. A., Lopez-Bigas, N., Ouzunis, C., Perez-Iratxeta, C., Andrade-Navarro, M. A., Adeyemo, A., Patti, M. E., Semple, C. A. M., and Hide, W. (2006). Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res.*, 34(doi:10.1093/nar/gkl381).

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.*, 98(9):5116–5121.

Yang, J. J. and Yang, M. C. (2006). An improved procedure for gene selection from microarray experiments using false discovery rate criterion. *BMC Bioinformatics*, 7:15.

EXPERIMENTS ON SOLVING MULTICLASS RECOGNITION TASKS IN THE BIOLOGICAL AND MEDICAL DOMAINS

Paolo Soda
Facoltà di Ingegneria, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo 21, Roma, Italy
p.soda@unicampus.it

Keywords: Statistical Pattern Recognition, Decomposition Methods, One-per-class, Reliability Estimation, Classifier Ensembles.

Abstract: Multiclass learning problems can be cast as the task of assigning instances to a finite set of classes. Although in the wide variety of learning tools there exist some algorithms capable of handling polychotomies, many of the tools were designed by nature for dichotomies. In the literature, many techniques that decompose a polychotomy into a series of dichotomies have been proposed. One of the possible approaches, known as *one-per-class*, is based on a pool of binary modules, where each one distinguishes the elements of one class from those of the others. In this framework, we propose a novel reconstruction criterion, i.e. a rule that sets the final decision on the basis of the single binary classifications. It looks at the quality of the current input and, more specifically, it is a function of the reliability of each classification act provided by the binary modules. The approach has been tested on four biological and medical datasets and the achieved performance has been compared with the one previously reported in the literature, showing that the method improves the accuracies so far.

1 INTRODUCTION

Many supervised pattern recognition tasks can be cast as the problem of assigning elements to a finite set of classes or categories. Such tasks are referred to as binary learning, or dichotomies, when they aim at distinguishing instances of two classes, whereas they are named multiclass learning, or polychotomies, if there are more categories.

There is a huge number of applications that require multiclass categorization. Some examples are text classification, object recognition and support to medical diagnosis, to name a few.

In the literature numerous learning algorithms have been devised for multiclass problems, such as neural networks or decision trees. However it exists a different approach that is based on the reduction of the multiclass task into multiple binary problems, referred to as *decomposition method*. The problem complexity is therefore reduced trough the decomposition of the polychotomy in less complex subtasks. The basic observation that supports such an approach is that in the literature most of the available algorithms, which handle classification problems, are best suited to learning binary function (Dietterich and Bakiri, 1995; Mayoraz and Moreira, 1997). Different dichotomizers, i.e. the discriminating functions that subdivide the input patterns in two separated classes, perform the corresponding recognition task. To provide the final classification, their outputs are combined according to a given rule, usually referred to as *decision* or *reconstruction rule*.

In the framework of decomposition methods for classification, the various methods proposed to-date can be traced back to the following three categories (Dietterich and Bakiri, 1995; Mayoraz and Moreira, 1997; Jelonek and Stefanowski, 1998; Masulli and Valentini, 2000; Allwein et al., 2001; Crammer and Singer, 2002; Hastie and Tibshirani, 1998; Kuncheva, 2005).

The first one, called *one-per-class*, is based on a pool of binary learning functions, where each one separates a single class from all the others. The assignment of a new input to a certain class can be performed, for example, looking at the function that returns the highest activation (Dietterich and Bakiri, 1995; Masulli and Valentini, 2000).

The second approach, commonly referred to as *distribuited output code*, assigns a unique codeword, i.e. a binary string, to each class. If we assume that the string has n bit, the recognition system is composed by n binary classification functions. Given an unknown pattern, the classifiers provide a n-bit string that is compared with the codeword to set the

final decision. For example, the input sample is assigned to the class with the closest codeword, according to a distance measure, such as the Hamming one. In this framework, in (Dietterich and Bakiri, 1995) the authors proposed an approach, known as *error-correcting techniques* (ECOC), where they employed error-correcting codes as a distributed output representation. Their strategy was a decomposition method based on the coding theory that allowed obtaining a recognition system less sensitive to noise via the implementation of an error-recovering capability. Although the traditional measure of diversity between the codewords and the outputs of dichotomizers is the Hamming distance, other works proposed different measures. For example, Kuncheva in (Kuncheva, 2005) presented a measure that accounted for the overall diversity in the ensemble of binary classifiers.

The last approach is called n^2 classifier. In this case the recognition system is composed of $(n^2-n)/2$ base dichotomizers, where each one is specialized in discriminating respective pair of decision classes. Then, their predictions are aggregated to a final decision using a voting criterion. For example, in (Jelonek and Stefanowski, 1998) the authors proposed a voting scheme adjusted by the credibilities of the base classifiers, which were calculated during the learning phase of the classification.

This short description of the methods so far shows that the recognition systems based on decomposition methods are constituted by an ensemble of binary discriminating functions. On this motivation, for brevity such systems are referred to as Multy Dichotomies System (MDS) in the following.

In the framework of the one-per-class approach, we present here a novel reconstruction rule that relies upon the quality of the input pattern and looks at the reliability of each classification act provided by the binary modules. Furthermore, the classification scheme that we propose allows employing either a single expert or an ensemble of classifiers internal to each module that solves a dichotomy. Finally, the effectiveness of the recognition system has been evaluated on four different datasets that belongs to biological and medical applications.

The rest of the paper is organized as follows: in the next section we introduce some notations and we present general considerations related to the system configuration. Section 3 details the reconstruction method and section 4 describes and discusses the experiments performed on four different medical datasets. Finally section 5 offers a conclusion.

2 PROBLEM DEFINITION

2.1 Background

Let us consider a classification task on c data classes, represented by the set of labels $\Omega = \{\omega_1, \cdots, \omega_c\}$, with $c > 2$. With reference to the one-per-class approach, the multiclass problem is reduced into c binary problems, each one addressed by one module of the pool $M = \{M_1, \cdots, M_c\}$. We say that the module, or the dichotomizer, M_j is specialized in the jth class when it aims at recognizing if the sample x belongs either to the jth class ω_j or, alternatively, to any other class ω_i, with $i \neq j$. Therefore each module assigns to the input pattern $x \in \Re^n$ a binary label:

$$M_j(x) = \begin{cases} 1 & \text{if } x \in \omega_j \\ 0 & \text{if } x \in \omega_i,\ i \neq j \end{cases} \quad (1)$$

where $M_j(x)$ indicates the output of the jth module on the pattern x. On this basis, the codeword associated to the class ω_j has a bit equal to 1 at the jth position, and 0 elsewhere.

Notice that we have just mentioned *module* and not *classifier* to emphasize that each dichotomy can be solved not only by a single expert, but also by an ensemble of classifiers. However, to our knowledge, the system dichotomizers typically adopt the former approach, i.e. they are composed by one classifier per specialized module. For example, for their experimental assessments the authors used a a decision tree and a multi layer perceptrons with one hidden layer both in (Mayoraz and Moreira, 1997) and (Masulli and Valentini, 2000), respectively. The same functions were employed by Dietterich and Bakiri for the evaluation of their proposal in (Dietterich and Bakiri, 1995), whereas Allwein et al. used a Support Vector Machine (Allwein et al., 2001). A viable alternative to using a single expert is the combination of classifiers outputs solving the same recognition task. The idea is that the classification performance attainable by their combination should be improved by taking advantage of the strength of the single classifiers. Classifier selection and fusion are the two main combination strategies reported in the literature. The former presumes that each classifier has expertise in some local area of the feature space (Woods et al., 1997; Kuncheva, 2002; Xu et al., 1992). For example, when an unknown pattern is submitted for classification, the more accurate classifier in the vicinity of the input is selected to label it (Woods et al., 1997). The latter algorithms assume that the classifiers are applied in parallel and their outputs are combined to attain somehow a group of "consensus" (De Stefano et al., 2000; Kuncheva et al., 2001; Kittler et al., 1998). Typi-

cal fusion techniques include weighted mean, voting, correlation, probability, etc..

It is worth noticing that the modules, besides labelling each pattern, may supply other information typically related to the degree that the sample belongs to that class. In this respect, the various classification algorithms are divided into three categories, on the basis of the output information that they are able to provide (Xu et al., 1992). The classifiers of *type 1* supply only the label of the presumed class and, therefore, they are also known as experts that work at the *abstract* level. *Type 2* classifiers work at the *rank* level, i.e. they rank all classes in a queue where the class at the top is the first choice. Learning functions of *type 3* operate at the *measurement* level, i.e. they attribute each class a value that measure the degree that the input sample belongs to that class. If a crisp label of the input pattern is needed, we can use the maximum membership rule that assigns x to the class for which the degree of support is maximum (ties are resolved arbitrarily). Although abstract classifiers provide a n-bit string that can be compared with the codewords, decision schemes that exploit information derived from the classifiers working at the measurement level permit us to define reconstruction rules that are potentially more effective. Furthermore, if the module is constituted by a multi-experts system, the information supplied by the single classifiers can be used to compute a measure similar to that provided by measurement classifiers.

Since measurement classifiers can provide more information with respect to the other two types, we assume that only measurement experts constitutes our MDS. Therefore, the research focus becomes: "Given the individual decision $M_1(x), \cdots, M_c(x)$ and the degrees of membership of x to the different classes, how can we use such an information to set the final label?".

2.2 The Reconstruction Method

The reconstruction method addresses the issues of determining the final label of the input pattern x on the basis of the modules' decisions and, eventually, of information directly derived from their outputs. To present our method, let us introduce two auxiliaries quantities. The first, named *binary profile*, represents the state of the module outputs. It is a c-bit vector defined by:

$$\mathbf{M}(x) = [M_1(x), \cdots, M_j(x), \cdots, M_c(x)] \qquad (2)$$

whose entries are the crisp labels provided by each module in the classification of sample x (see equation 1).

Since each block has a binary output, the 2^c possible bit combinations of the binary profile can be grouped into the following three categories:

(i) only one module classifies the sample in the class in which it is specialized;

(ii) more modules classify the sample in its own class;

(iii) none module classifies the sample in its own class.

In the first case, only one entry of $\mathbf{M}(x)$ is one; in the second more elements are one (at least two and no more than c), whereas in the last situation all the elements are zero. Such an observation naturally leads to distinguish these three cases using the summation over the binary profile. Indeed,

$$m = \sum_{j=1}^{c} M_j(x) = \begin{cases} 1, & \text{in case (i)} \\ [2,c], & \text{in case (ii)} \\ 0, & \text{in case (iii)} \end{cases} \qquad (3)$$

where m therefore represents the number of modules whose outputs are 1.

The second quantity that we introduce is referred to as *reliability profile* and it is described by:

$$\psi(x) = [\psi_1(x), \cdots, \psi_j(x), \cdots, \psi_c(x)] \qquad (4)$$

where each element $\psi_j(x)$ measure the reliability of the classification act on pattern x provided by the jth module. Note that the reliability varies in the interval $[0,1]$, and a value near 1 indicates a very reliable classification.

We deem that the estimation of the reliability of each classification act is a viable method to employ the information directly derived from the classifiers output since it has demonstrated its convenience, in other field also (De Stefano et al., 2000; Cordella et al., 1999).

Assuming that we determined both the binary and the reliability profiles, i.e. $\mathbf{M}(x)$ and $\psi(x)$ respectively, in the next section we will present the reconstruction rule.

3 RELIABILITY BASED RECONSTRUCTION

In this section we introduce the novel reconstruction strategy we propose in the paper. It chooses an output in any of the 2^c combinations of the binary profile. We deem that an accurate final decision can be taken if the reconstruction rule looks at the quality of the classification provided by the modules, i.e. at the reliability of their specific decisions. To our knowledge the application of such a parameter can not be found in the literature related to decomposition methods. Indeed, the papers of this field that used the information

directly derived from the outputs of the base classifiers typically considered only the highest activation among the experts, e.g. the maximum output from a pool of neural networks. However, this measure cannot be regarded as a reliability parameter, since it has been demonstrated that it should be computed considering not only the winner output neurons but also the losers (Cordella et al., 1999).

Therefore, differently from the past, we propose a criterion that makes use of the reliability measure, i.e. of the reliability profile, named as *Reliability-based-Reconstruction (RbR)*. Denoting by s the index of the module that sets the final output $O(x) \in \Omega$, referred to as selected module for brevity in the following, the final decision is given by:

$$O(x) = \omega_s \tag{5}$$

with

$$s = \begin{cases} \arg\max_j(M_j(x) \cdot \psi_j(x)), \text{ if } m \in [1,c] \\ \arg\min_j(\overline{M_j(x)} \cdot \psi_j(x)), \text{ if } m = 0 \end{cases} \tag{6}$$

where $\overline{M_j(x)}$ indicates the negate output of the jth block.

The first row of this equation considers both cases (i) and (ii). Indeed, since in the first case all the modules agree in their decision, as a final output is chosen the class of the module whose output is 1. Conversely, in cases (ii) and (iii) the final decision is performed looking at the reliability of each modules' classifications. In case (ii), m modules vote for their own class, whereas the others $(c-m)$ indicate that x does not belong to their own class. To solve the dichotomy between the m conflicting modules we look at the reliability of their classification and choose the class associated to the more reliable one. In case (iii) $m = 0$, suggesting that all modules classify x as belonging to another class than the one they are specialized. In this case, the bigger is the reliability parameter $\psi_j(x)$, the less is the probability that x belongs to ω_j, and the bigger is the probability that it belongs to the other classes. These observations suggest finding out which module has the minimum reliability and then choosing the class associated to it as a final output.

Panel *A* of figure 1 shows the architecture of the proposed recognition system. The decision $M_j(x)$ and the reliability $\psi_j(x)$ supplied by each of the c modules are aggregated in the *reconstruction module* to provide the final decision $O(x)$. As observed in section 2.1, the use of an ensemble of classifiers in each module is a way to improve its discrimination capability. In this respect, the panel *B* of the same figure depicts a typical configuration of a multi-experts system. Notice that both the output of the kth classifier and its reliability, denoted as $V_k(x)$ and $\xi_k(x)$, respectively, can be given to the combination rule in order to label the input sample.

4 EXPERIMENTAL EVALUATION

In this section we first describe the datasets used to assess the performance of the reconstruction method and, second, we briefly discuss the configuration of the MDS modules. Third, we present a strategy to evaluate the classification reliability when the modules are constituted both by a single classifier and by an ensemble of experts, respectively. Finally, we report the experimental results.

4.1 Datasets

For our tests we use four datasets, described in the following and summerized in table 1.

Indirect Immunofluorescence Well Fluorescence Intensity. Connective tissue diseases are autoimmune disorders characterized by a chronic inflammatory process involving connective tissues. When they are suspected in a patient, the Indirect Immunofluorescence (IIF) test based on HEp-2 substrate is usually performed, since it is the recommended method. The interested reader may find a wide explanation of the IIF and its issues in (Kavanaugh et al., 2000; Rigon et al., 2007). The dataset consists of 14 features extracted from 600 patients sera collected at Università Campus Bio-Medico di Roma. The samples are distributed over three classes, namely positive (36.0%), negative (32.5%) and intermediate (31.5%). Previous results are reported in (Soda and Iannello, 2006) where the authors employed a multiclass approach, achieving an accuracy of 76% approximately.

Indirect Immunofluorescence HEp-2 cells staining pattern. This is a dataset with 573 instances represented by 159 statistical and spectral features. The samples are distributed in five classes that are representative of the main staining patterns exhibited by HEp-2 cells, namely homogeneous (23.9%), peripheral nuclear (21.8%), speckled (37.0%), nucleolar (8.2%) and artefact (9.1%). These patterns are related to one of the different autoantibodies that give rise to a connective tissue disease. For the details on these classes see (Rigon et al., 2007). On this dataset, we performed some tests adopting a multiclass approach, which exhibits a hit rate of 63.6% approximately, evaluated using the eightfold cross validation.

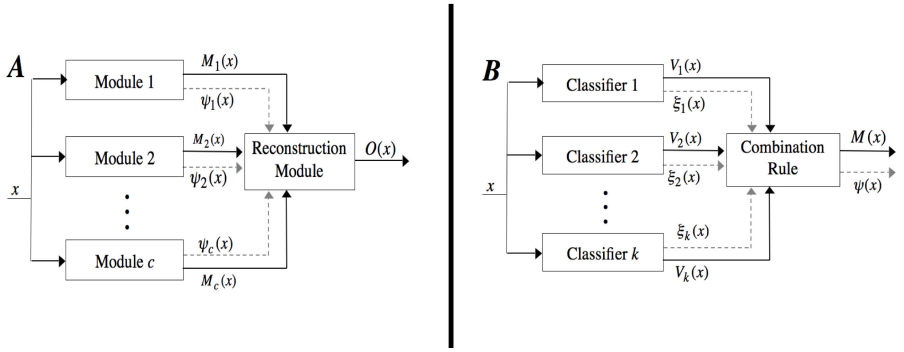


Figure 1: The system architecture, which is based on the aggregation of binary modules (panel ***A***), according to the one-per-class approach. Note that each module can be constituted by a multi-experts system, as depicted in the panel ***B***.

Lymphography. A database of lymph diseases was obtained from the University Medical Centre, Institute of Oncology, Ljubljana. It is composed by 148 instances described by 18 numeric attributes. There are four classes, namely normal (1.4%), metastases (54.7%), malign lymph (41.2%) and fibrosis (2.7%). The data are available within the UCI Machine Learning Repository[1] (Asuncion and Newman, 2007). Different approaches were used in the literature to address the recognition task. For instance, for Naive Bayes classifier and C4.5 decision tree the achieved performance was 79% and 77% respectively (Clark and Niblett, 1987), whereas induction techniques correctly classified the 83% of samples (Cheung, 2001).

Ecoli. The database is composed by 336 samples, described by a nine-dimensional vector and distributed in eight classes. Each class represents a localization site, which can be cytoplasm (42.5%), inner membrane without signal sequence (22.9%), periplasm (15.5%), inner membrane, uncleavable signal sequence (10.4%), outer membrane (6.0%), outer membrane lipoprotein (1.5%), inner membrane lipoprotein (0.6%) and inner membrane, cleavable signal sequence (0.6%). Again, the data are available within the UCI Machine Learning Repository (Asuncion and Newman, 2007). In (Jelonek and Stefanowski, 1998), the authors reported an accuracy that ranges from 79.7% up to 83.0%, achieved employing both a decision tree and a Multi Layer Perceptrons, respectively. In (Allwein et al., 2001), using many popular classification algorithms, such as the support-vector machines, AdaBoost, regression and decision-tree algorithms, the hit rate varies from 78.5% up to 86.1%.

[1]For each dataset of this repository the users have access to a description of the application domain, to the features and to the ground truth.

4.2 MDS Configuration

The modules of the MDS are essentially composed by a single classifier or by an ensemble of classifiers. In both cases, as single expert we use k-Nearest Neighbour (kNN) or Multi-Layer Perceptron (MLP). For each dichotomy, we first select a subset of features that simplifies both the pattern representation and the classifier complexity as well as the risk of the incurring in the peaking phenomenon[2]. Then we carry out some preliminary tests to determine the best configuration of experts parameters, e.g. the number of neighbours for kNN classifier or the number of hidden layers, neurons per layer, etc., for the MLP network. Furthermore, when the module is constituted by an ensemble of experts we adopt a fusion technique to combine their outputs, namely the Weighted Voting (WV). In such a method the opinion of each expert about the class of the input pattern is weighted by the reliability of its classification. Since each expert deals with a binary learning task, to further present this scheme we can simplify the notation as follows. Denoting as $V_k(x)$ and as $\xi_k(x)$ the output and the classification reliability of kth classifier on sample x, the weighted sum of the votes for each of the two classes is given by:

$$W_h(x) = \sum_{k:V_k(x)=h} \xi_k(x), \text{ with } h = \{0,1\} \qquad (7)$$

[2]The peaking phenomenon is a paradoxical behaviour in which the added features may actually degrade the performance of a classifier if the number of training samples that are used to design the classifier is small relative to the number of features.

Table 1: Summary of the datasets used.

Database	Number of Samples	Number of Classes	Number of features	Avalaibility
IIF Well Fluorescence Intensity	600	3	14	Private
IIF HEp-2 cells staining pattern	573	5	159	Private
Lymphography	148	4	18	UCI
Ecoli	336	8	9	UCI

The output of the fusion of the jth module, $M_j(x)$, is defined by[3]:

$$M_j(x) = \begin{cases} 1 & \text{if } W_1(x) > W_0(x) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Turning our attention to the configuration of the system in the experimental tests, notice that the modules that label the samples of the IIF Well Fluorescence Intensity and of lymphography datasets are composed by one classifiers. The modules that classify the samples of the HEp-2 cells and of the ecoli databases are constituted by kNN and MLP classifiers combined by the WV criterion.

4.3 Reliability Parameter

The approach described for deriving the final decision according to the RbR rule requires the introduction of a reliability parameter that evaluates the quality of the classification performed by each module, which can be composed by a single classifier or by an aggregation of classifiers (figure 1). In the former case its reliability ψ_j coincides with the one of the single classifier, i.e. ξ. In the latter case, each entry of the reliability profile generally depends on the combination rule adopted in the module, on the number k of composing experts and on their individual reliabilities ξ. Formally,

$$\psi_j(x) = \begin{cases} \xi(x), & \text{if } k = 1 \\ f(\xi_1(x), \cdots, \xi_i(x), \cdots, \xi_k(x)), & \text{if } k > 1 \end{cases} \quad (9)$$

where all the reliabilities are reported as function of the input pattern to emphasize that they are computed for each classification act.

In the rest of this section we first present two techniques to measure the reliability of kNN and MLP decisions, and then we introduce a novel method that estimates such parameter in the case of the application of the WV criterion.

A typical approach that measures the reliability of the decision taken by the single expert, i.e. ξ, makes use of the confusion matrix[4] estimated on the learning set. The drawback of this method is that all the patterns with the same label have equal reliability, regardless of the quality of the sample. Indeed, the average performance on the learning set, although significant, does not necessarily reflect the actual reliability of each classification act. To overcome such limitations we adopt an approach that relies upon the quality of the current input. To this end, we refer to the work presented in (Cordella et al., 1999), where the quality of the sample is related to its position in the feature space. In this respect, the low reliability of a recognition act can be traced back to one of the following situations: (a) in the feature space x is located in a region that is far from those associated with the various classes, i.e. the sample is significantly different from those present in the training set, (b) the point representing x lies in a region of the feature space where two or more classes overlap. These observations lead to introduce the parameters ξ_a and ξ_b that distinguish between the two situations of unreliable classification. Then, a comprehensive parameter ξ can be derived adopting the following conservative choice:

$$\xi = min(\xi_a, \xi_b) \quad (10)$$

Indeed, it implies that a low value for only one of the parameters is sufficient to consider unreliable the classification.

In the case of kNN classifiers, following (Cordella et al., 1999), the two parameters are defined are given by:

$$\xi_a = \max(1 - D_{min}/D_{max}, 0) \quad (11)$$

$$\xi_b = 1 - D_{min}/D_{min2} \quad (12)$$

where D_{min} is the smallest distance of x from a reference sample belonging to the same class of x, D_{max} is the highest among the values of D_{min} obtained for samples taken from the training-test set, i.e. a set that is disjoint from both the reference and the test set, D_{min2} is the distance between x and the reference sample with the second smallest distance from x among

[3] In case of tie, i.e. if $W_1(x)$ is equal to $W_0(x)$, the output $M_j(x)$ is set arbitrarily to zero. Note that it never occurred in all tests we performed.

[4] The confusion matrix reports for each entry (p,q) the percentage of samples of the class C_p assigned to the class C_q.

all the reference set samples belonging to a class that is different from that determining D_{min}.

In the case of MLP classifier, the two quantities are defined as follows:

$$\xi_a = N_{win} \tag{13}$$

$$\xi_b = N_{win} - N_{2win} \tag{14}$$

where N_{win} is the output of the winner neuron, N_{2win} is the output of the neuron with the highest value after the winner. From this definition, it is straightforward that $\xi = \xi_b$. For further details see (De Stefano et al., 2000).

When the jth module is composed by more than one classifier combined according to the WV rule, the reliability estimator considers again the situations which can give rise to an unreliable classification. In this respect, we need to introduce the following two auxiliary quantities:

$$\pi_1(x) = \max\left(\{\xi_k(x) | k : V_k(x) = M_j(x)\}\right) \tag{15}$$

$$\pi_2(x) = \max\left(\{\xi_k(x) | k : V_k(x) \neq M_j(x)\} \cup \{0\}\right) \tag{16}$$

where $\pi_1(x)$ and $\pi_2(x)$ represent the maximum reliabilities of experts voting for the winning class and for other classes (0 if all the experts agree on the winner class), respectively. Given these definitions, the reliability of the WV rule can be evaluated according to the following conservative choice:

$$\psi(x) = \min\left(\pi_1(x), \max\left(0, 1 - \pi_2(x)/\pi_1(x)\right)\right) \tag{17}$$

4.4 Results and Discussion

This section presents the experimental results that we achieved using the system described so far. To evaluate and then compare the results of this approach with those reported in the literature we perform eightfold and tenfold cross validation on the two IIF datasets, i.e. well fluorescence intensity and HEp-2 cells staining pattern, and on the other two databases, i.e. lymphography and ecoli, respectively.

The third column of table 2 shows the testing accuracies achieved on the four databases. To simply compare them with the past results, the second column of the same table summarizes the performance reported in literature. Turning our attention to the tests carried out on the first and on the second datasets, a relevant accuracy improvement can be observed. Indeed, the hit rate increases of 18.4% and of 12.3% in the case of the well fluorescence intensity and HEp-2 cells staining pattern databases, respectively. In our opinion, such an improvement is twofold motivated. On the one hand, the set of extracted features is more stable and more effective when we adopt a decomposition approach rather than a multiclass one. On the other hand, the reconstruction rule exhibits a very good capability of solving the disagreements between the specialized modules. Indeed, when the binary profile of the input sample $\mathbf{M}(x)$ differs from one of the possible codewords (i.e. $m = 0$ or $2 \leq m \leq c$), the decision is taken looking at the reliability profile $\psi(x)$, as presented in the formula 6. These considerations are strengthened by the observation of the performance attained in the classification of samples belonging to the two UCI datasets. Indeed, since they are benchmark datasets, any reported improvement is due to the recognition approach rather than to the use of a different features set. The tests on both the lymphography and ecoli datasets exhibit an accuracy better than the one reported to date. Indeed, for the former dataset the improvement ranges both from 6.9% up to 12.9% , whereas for the latter one it varies from 1.8% up to 9.4%. Therefore, also in these cases the MDS in combination with the RbR rule improves the recognition performance. Furthermore, it is worth noting that the approach seems independent of the modules' arrangement. The rationale lies in observing that in two of the four tests the MDS modules are constituted by a multi-experts system, whereas in the others they are composed by a single classifier (see the beginning of section 4). Consequently, the reliability ψ_j is measured according to a method that varies with the module configuration, as previously presented (see equations 10-17). Nevertheless, these variations do not affect the effectiveness of the recognition system. Therefore, we deem that the reconstruction rule is robust with respect to different reliability estimators.

5 CONCLUSIONS

In the framework of decomposition methods, we have presented a classification approach that reconstructs the final decision looking at the reliability of each classification act provided by all dichotomizers. Furthermore, the reconstruction rule does not depend on the configuration of each module, i.e. on its architecture. Such an observation is strengthened by the good performance achieved when both a single classifier and a fusion of experts constitute each module, respectively.

For all the four tested databases, the experimental results show that the proposed system outperforms the performance reported in the literature.

Future works are directed towards two issues. First, the test of the system on other public datasets and, second, the definition of reliability parameter of

Table 2: Testing accuracy achieved on the used datasets.

Database	Past Usage	MDS using RbR
IIF Well Fluorescence Intensity	75.9%	94.3%
IIF HEp-2 cells staining pattern	63.6%	75.9%
Lymphography	$77\% - 83.0\%$	89.9%
Ecoli	$78.5\% - 86.1\%$	87.9%

each decision taken by the MDS.

ACKNOWLEDGEMENTS

The author would like to thank the DAS s.r.l (`www.dasitaly.com`), which has funded this work.

REFERENCES

Allwein, E. L., Schapire, R. E., and Singer, Y. (2001). Reducing multiclass to binary: a unifying approach for margin classifiers. *J. Mach. Learn. Res.*, 1:113–141.

Asuncion, A. and Newman, D. J. (2007). UCI machine learning repository.

Cheung, N. (2001). Machine learning techniques for medical analysis. Master's thesis, University of Queensland.

Clark, P. and Niblett, T. (1987). Induction in noisy domains. In *Progress in Machine Learning–Proc. of EWSL 87*, pages 11–30.

Cordella, L., Foggia, P., and et. al. (1999). Reliability parameters to improve combination strategies in multi-expert systems. *Patt. An. & Appl.*, 2(3):205–214.

Crammer, K. and Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2:265–292.

De Stefano, C., Sansone, C., and Vento, M. (2000). To reject or not to reject: that is the question: an answer in case of neural classifiers. *IEEE Trans. on Systems, Man, and Cybernetics–Part C*, 30(1):84–93.

Dietterich, T. G. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263.

Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. In *NIPS '97: Proc. of the 1997 Conf. on Advances in neural information processing systems*, pages 507–513. MIT Press.

Jelonek, J. and Stefanowski, J. (1998). Experiments on solving multiclass learning problems by n^2 classifier. In *10th European Conference on Machine Learning*, pages 172–177. Springer-Verlag Lecture Notes in Artificial Intelligence.

Kavanaugh, A., Tomar, R., and et al. (2000). Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. *Am. Col. of Pathologists, Archives of Pathology and Lab. Medicine*, 124(1):71–81.

Kittler, J., Hatef, M., and et. al. (1998). On combining classifiers. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 20(3):226–239.

Kuncheva, L. I. (2002). Switching between selection and fusion in combining classifiers: an experiment. *IEEE Trans. on Systems, Man and Cybernetics, Part B*, 32(2):146–156.

Kuncheva, L. I. (2005). Using diversity measures for generating error-correcting output codes in classifier ensembles. *Patt. Recogn. Lett.*, 26(1):83–90.

Kuncheva, L. I., Bezdek, J. C., and Duin, R. (2001). Decision template for multiple classifier fusion: an experimental comparison. *Patt. Recognition*, 34:299–314.

Masulli, F. and Valentini, G. (2000). Comparing decomposition methods for classication. In *KES'2000, Fourth Int. Conf. on Knowledge-Based Intell. Eng. Systems & Allied Technologies*, pages 788–791.

Mayoraz, E. and Moreira, M. (1997). On the decomposition of polychotomies into dichotomies. In *ICML '97: Proc. of the 14th Int. Conf. on Machine Learning*, pages 219–226. Morgan Kaufmann Publishers Inc.

Rigon, A., Soda, P., Zennaro, D., Iannello, G., and Afeltra, A. (2007). Indirect immunofluorescence (IIF) in autoimmune diseases: Assessment of digital images for diagnostic purpose. *Cytometry - Accepted for Publication, February*.

Soda, P. and Iannello, G. (2006). A multi-expert system to classify fluorescent intensity in antinuclear autoantibodies testing. In *Computer Based Medical Systems*, pages 219–224. IEEE Computer Society.

Woods, K., Kegelmeyer, W., and Bowyer, K. (1997). Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:405–410.

Xu, L., Krzyzak, A., and Suen, C. (1992). Method of combining multiple classifiers and their application to handwritten numeral recognition. *IEEE Trans. on Systems, Man and Cybernetics*, 22(3):418–435.

IMAGE SEGMENTATION TO EVALUATE ISLETS OF LANGHERANS

C. Grimaudo, D. Tegolo, C. Valenti
Dipartimento di Matematica e Applicazioni, Università di Palermo, via Archirafi 34, 90123, Italy
{*tegolo,cvalenti*}*@math.unipa.it*

F. Bertuzzi
Istituto Mediterraneo per i Trapianti e Terapie ad Alta Specializzazione, via Tricomi 1, 90127, Italy
fbertuzzi@ismett.edu

Keywords: Islets of Langherans, implantation advisability, image segmentation.

Abstract: This contribution deals with an unsupervised system to process digital photomicrographs in order to locate and analyze islets of Langherans in human pancreases. The experiment has been conducted on real data and, though we are still going to complete the evaluation of the whole method, we expect to define a set of proper features (e.g. area, perimeter, fractal dimension, shape complexity, texture and entropy) useful for a fast and reliable counting of healthy cells. In particular, this research aims to measure the advisability of a possible implantation in patients affected by type 1 diabetes mellitus.

1 INTRODUCTION

This paper introduces a new system for the automatic analysis of high power magnification photomicrographs of the human islets of Langherans. The cells that make these clusters can be divided into a few classes which include the α cells, that secrete glucagon, and the β cells, responsible for the production of insulin. This research field is of particular interest because of the demand to evaluate the state of these endocrine tissue for preoperative planning in patients that suffer from severe type 1 diabetes mellitus, otherwise scarcely treatable by conventional therapies (Ryan et al., 2005; Shapiro et al., 2006). It has been verified that the probability of obtaining a favorable implantation increases when a large number of viable and purified islets is transplanted in to the patients (Bertuzzi and Ricordi, 2007). In a multivariate analysis aimed to identify some *in vitro* parameters for islet quality or function predictive of *in vivo* graft function of the same islets after their transplantation in diabetic patients, islet morphology (in terms of the maintenance of their round shape profile, similar to what they showed in the native pancreas) was demonstrated to be correlated with 1 month recipient c-peptide production (Ricordi et al., 2001); islet morphology therefore should be considered an indirect parameter of islet viability. These results call for the identification of some standardized strategies to characterize islet morphology and to quantify their degree of maintenance of their native round morphology (Nano et al., 2005).

At present, the analysis is also performed by improving the appearance through image processing softwares or ad hoc systems (Metamorph). A grid is laid on the slide so to fix the islets and to let easily count their different typologies (see Figure 1). This process is done by hand to separate those cells useful to the implantation and obviously it is slow, subjective and liable to errors; an environment to help the expert analyst is therefore desirable both to enhance the quality of the digital photos and to elaborate the images in order to locate automatically the zones of interest.

A variety of methods is already present in literature for both supervised and unsupervised segmentation of photomicrographs depicting cells (Coelho et al., 2002; Tripodo et al., 2006; Montseny et al., 2004; Bak et al., 2004). Usually these techniques are taken back to the elaboration of histograms, application of mathematical morphology, texture analysis, Fourier and wavelet transforms to extract the shapes of the components that have been found. Often the images have noise due to the presence of small artifacts, distortions and blurring introduced by the optical system, inherent inaccuracies due to the lattice

(e.g. the thickness of the sample that must be analyzed), imperfections of the coloring of the contrast agent (e.g. due to variations of exposure time and to the quantity of the marker itself).

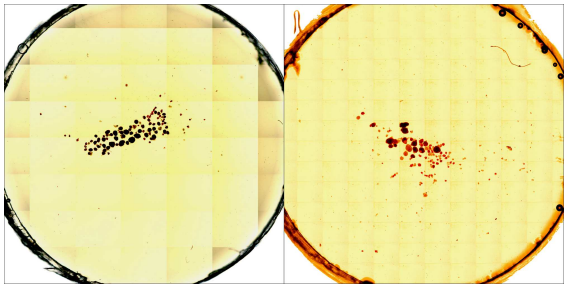


Figure 1: Two slides that put in evidence the presence of a grid to easily count the number of the islets.

We have realized a system that acts through a library of tools to preprocess the data; the segmentation of the various components in the images often requires the intervention of an expert user who locates the promising clusters of cells. This approach can be applied not only to the islets of Langherans, but also to analyze other vital cells (e.g. hepatocytes, bone marrow cells). Finally it should be verified the possibility to apply this strategy also in fixed tissue after different immunohystochemistry staining.

The following Section 2 describes the new environment to elaborate and classify the islets of Langherans. Experimental results are presented in Section 3, while remarks and possible future works are introduced in Section 4.

2 SEGMENTATION OF THE ISLETS

In this paper, we aimed to describe the system which has been developed to provide an unsupervised analysis of the human islets of Langherans (see Figure 2). Different techniques have been implemented to enhance the quality of the images, to segment all components, to distinguish among the cells and to evaluate their conditions in order to quantify the advisability of the implantation.

The photos in our database have been acquired through a digital tool; they suffer from artifacts due to the equipment (e.g. only the center of the image is correctly in focus and a few impurities can be present on the lenses). Predetermined threshold values result in a poor separation between the components of the images, but we have experimentally verified that the Otsu method (Otsu, 1979) is able to compute these optimal values in order to locate the imperfections on the red and green channels of the RGB color space. We have carried out a statistical examination on both the background and foreground to determine their starting threshold values; should the input image be very different from the database we have considered, then, to better calibrate the values, the user can select some regions of interest, representative of the different parts of the islets. Figure 3 shows the previous input image soon after the preprocessing step.

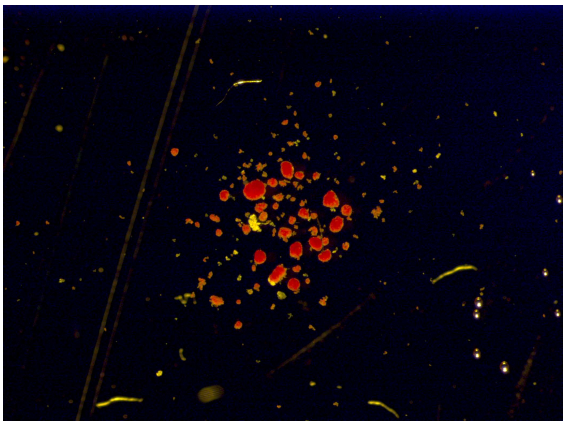


Figure 2: A sample photomicrograph of a cluster of the human islets of Langherans.

We have successfully applied the same adaptive self-tuning technique that has been introduced in (Tripodo et al., 2006) to discriminate between the pure β cells, or the mixed β and exocrine/ductal cells that are highlighted by the marker as red and orange zones respectively, while the yellow parts correspond to dead cells or impurities or simply exocrine/ductal cells. This usually leads to a rough representation of the cells, but a simple median filter is sufficient to remove all small objects (5×5 kernel) and pointlike noise (3×3 kernel). The shape of the cells so far obtained can be further enhanced by the use of a mathematical morphology opening with a structuring element represented by discrete disk of radius 2 (Soille, 2003). In such a way the cells of the islets are better separated and, moreover, we can safely delete all components that are too small (the allowed number of pixels has been pre-defined according to the present magnification power of the microscope). Figure 4 shows the final result obtained on the reference image; another example is reported in Figure 5. We have highlighted the final contour just to easily check the segmentation of the relevant islets.

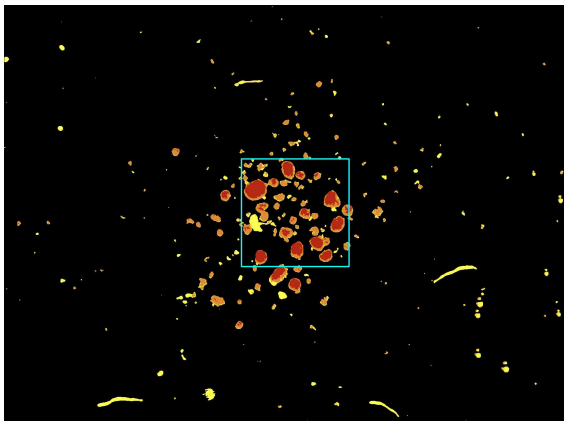


Figure 3: Some artifacts present in Figure 2 have been removed. Due to the huge field of view, in the following we will propose the results relative to the superimposed box.

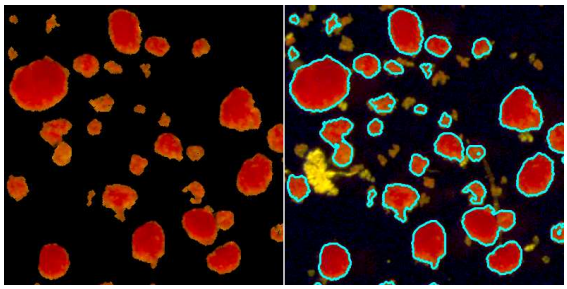


Figure 4: Left: the remaining artifacts and dead cells have been automatically removed from Figure 3. Right: the final contour has been plotted on the input image of Figure 2.

3 EXPERIMENTAL RESULTS

Images have a size of 2088×1550 pixels and were acquired at a sample dilution equal to 2500×, by a stereomicroscope Leica MZ12-5 with a 2× zoom magnification and equipped with a digital camera, able of a 4.34765μm/pixel picture calibration. The set of images we have studied has been obtained by isolating the islets through the automated method from multi-organ donors (Ricordi et al., 1989). After pancreas digestion the islets from 3 preparations have been purified by COBE processor (Vargas et al., 1996) and placed in a culture media for additional 48 hours at 24°C. The islets have been finally stained with dithizone (a vital stain that cross-reacts with zinc) and therefore it has been used to recognize the α, exocrine and ductal cells (in which zinc is absent) from the β cells (rich in zinc).

A set of parameters that describe each kind of cluster of cells has been extracted from the segmented images. The area, the perimeter, the compactness (i.e. the normalized ratio between the area and the squared perimeter) and the eccentricity of the ellipse which approximates the shape of the islet and the measures of convexity/concavity of its edges return a quantitative esteem of its aspect. In particular, compactness and eccentricity measure the roundness: healthy islets should not have protrusions.

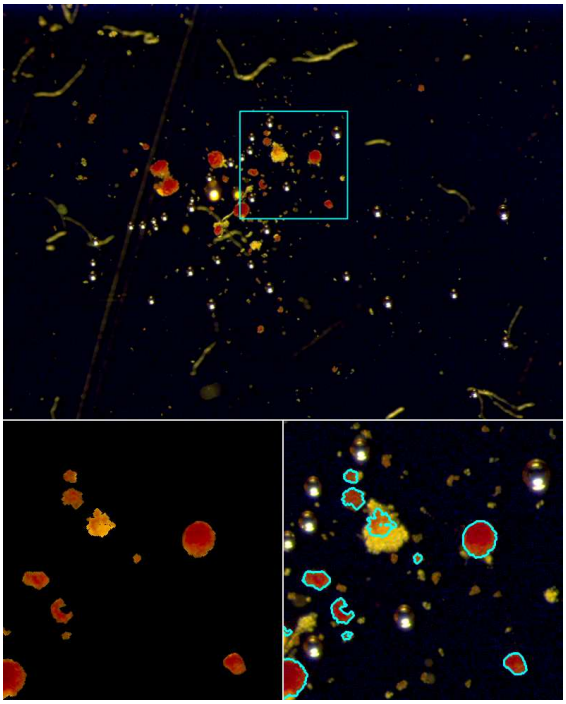
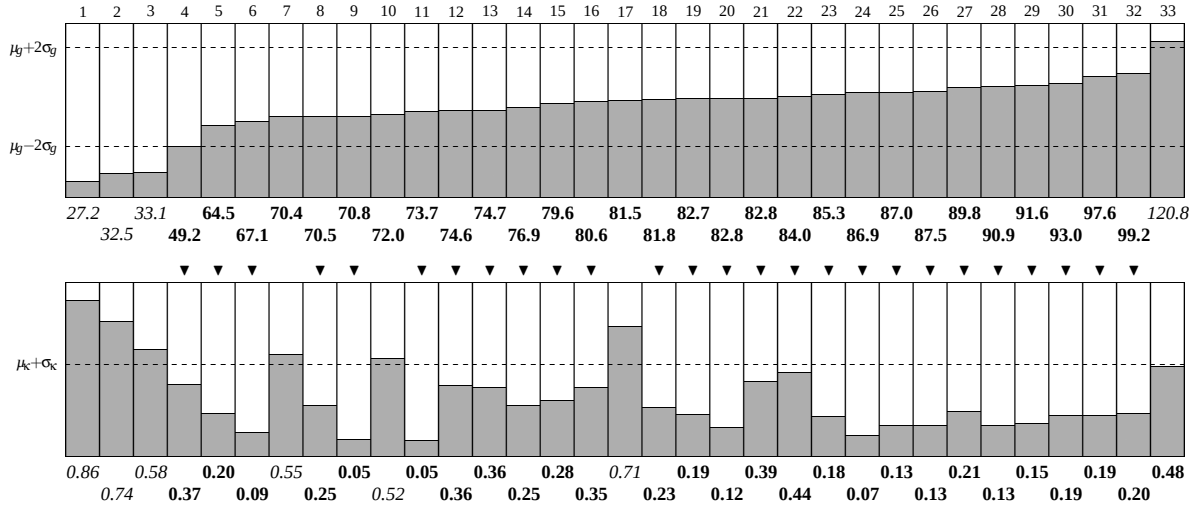


Figure 5: The edges of the islets within the box have been marked in blue.

The amount of information directly deducible from the luminosity of the pixels is another useful characteristic: the more homogenous an islet is, the smaller its local entropy is. We are still investigating on the ability of operators that return marks about the value of local sharpness and textures (which are closely connected to the presence of luminosity gradients).

For each islet $\imath_i$ we compute the product g_i between its average luminosity ℓ_i and its entropy e_i. If we indicate with μ_g and σ_g respectively the mean and the standard deviation of all $g = \ell \times e$, then the islets with a score $|g_i - \mu_g| < 2\sigma_g$ can be considered as promising candidate. A further important characteristic is given by the compactness κ (Rangayyan, 2005): with an analogous approach, the islets till now accepted with a compactness value $\kappa_i < \mu_\kappa + \sigma_\kappa$ are definitely classified as reasonably good. For the sake of clarity, an islet is classified as good if it passes the test on g and then on κ. The final evaluation of the whole input photo of Figure 5 is summarized in Ta-

Table 1: The features of each islet (33 in this example) have been represented by two columns (top: g and bottom: κ). The threshold values are represented by dashed lines and both tests have to be passed: κ reduces the number of candidates already obtained by g (good islets have been marked with a ▼).



ble 1. Figure 6 shows how the system highlights a single islet and proposes its features.

The percentage of the area of the yellow zones (more precisely, the ratio between red and yellow) indicates the purification of islet preparation and the eventual presence of embedded islets, that means islets surrounded by exocrine tissue (Ricordi et al., 1995). The final ratio between the area of good islets and the area of all islets summarizes the goodness of the inspected photomicrograph. Several parameters have been therefore available now by an automated method of analysis for the characterization of an islet preparation in terms of:

- islet number (the number of red clusters);
- islet dimension (the red area);
- islet purification (the ratio between yellow and red areas in the whole preparation);
- percentage of embedded islets (the ratio between red and yellow areas within an islet);
- islet morphology.

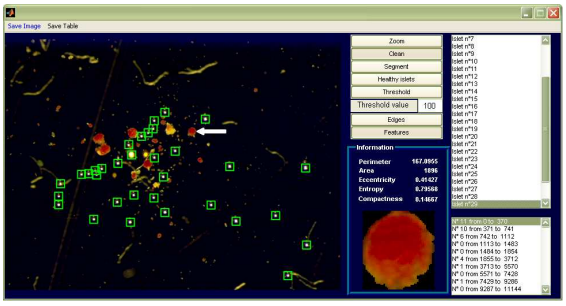


Figure 6: A screenshot of the graphical interface of the system. Small green boxes automatically delimit bubbles (present as artifacts in the photo). A selected islet is pointed out by a white arrow and the values of the relevant features are presented to the user.

4 REMARKS AND FURTHER WORKS

We have introduced an unsupervised system to locate the human islets of Langherans in photomicrographs. These clusters of cells have been characterized in order to define some parameters representative of their number and morphology. The predictive role of these features towards their *in vivo* graft function should be matter of further studies.

From a computer science point of view, the efficiency of the proposed method is still at the testing stage (Altman, 1999) and our system should be considered as a tool to help the experts in obtaining a quantitative esteem of the reliability of the islets in favorable implantation. The final results have been validated by biologists involved in implantations to treat patients affected by severe forms of type 1 diabetes mellitus. It is interesting to note that the methodologies we have applied to segment the components of the photos are quite standard and general enough and that the extracted features can be extended to differentiate between the α and β cells which compose the islets; this is to correlate their peculiarities with information of the state of the patients. Moreover, though preliminary results are encouraging, we are improving the segmentation procedure by including further algorithms based on mathematical morphology and

watershed/level sets.

To the best of our knowledge, our environment is the first attempt to automatically analyze islets of Langherans for implantations. Previous works rely on manual segmentation of their photomicrographs or are too general, thus to require to be adapted in order to process images containing these kind of cells. Therefore, a comparison of the results obtained by our system is still desirable.

Additional projects should be the *in vitro* characterization of the human islet preparations after the staining with vital probes (i.e. propidium iodide, fluorescein diacetate (Barnett et al., 2004; Miyamoto et al., 2000) and probes for apoptosis (Ichii et al., 2005). This should allow the direct quantification of vital, apoptotic and necrotic islets. Finally the automated system for imaging analysis should be applied in fixed tissues after immunostaining for insulin and glucagon thus allowing a complete characterization of islet cell composition (Ichii et al., 2005; Street et al., 2004).

ACKNOWLEDGEMENTS

The authors wish to thank Doctor Domenico Bosco of the Hôpitaux Universitaires de Genève for useful discussions and his kind contribution in providing the input images.

REFERENCES

D.Altman (1999). *Practical statistics for medical research*. Chapman & Hall/CRC.

E.Bak, K.Najarian, J.P.Brockway (2004). Efficient segmentation framework of cell images in noise environments. *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society*, 3:1802–1805.

M.J.Barnett, D.McGhee-Wilson, A.M.Shapiro, J.R.Lakey (2004). Variation in human islet viability based on different membrane integrity stains. *Cell Transplant*, 13:481–488.

F.Bertuzzi, C.Ricordi (2007). Prediction of clinical outcome in islet allotransplantation. *Diabetes Care*, 30:410–417.

R.C.Coelho, V.Gesù, G.Bosco, J.S.Tanaka, C.Valenti (2002). Shape-based features for cat ganglion retinal cells classification. *Real-Time Imaging, Special Issue on Imaging in Bioinformatics*, 8:213–226.

H.Ichii, L.Inverardi, A.Pileggi, R.D.Molano, O.Cabrera, A.Caicedo, S.Messinger, Y.Kuroda, P.O.Berggren, C.Ricordi (2005). A novel method for the assessment of cellular composition and beta-cell viability in human islet preparations. *Am J Transplant*, 5:1635–1645.

M.Miyamoto, Y.Morimoto, Y.Nozawa, A.N.Balamurugan, B.Xu, (2000). Establishment of fluorescein diacetate and ethidium bromide (fdaeb) assay for quality assessment of isolated islets. *Cell Transplant*, 9:681–686.

E.Montseny, P.Sobrevilla, S.Romani (2004). A fuzzy approach to white blood cells segmentation in color bone marrow images. *Proceedings of the IEEE International Conference on Fuzzy Systems*, 1:173–178.

R.Nano, B.Clissi, R.Melzi, G.Calori, P.Maffi, B.Antonioli, S.Marzorati, L.Aldrighetti, M.Freschi, T.Grochowiecki, C.Socci, A.Secchi, V.Carlo, E.Bonifacio, F.Bertuzzi (2005). Islet isolation for allotransplantation: variables associated with successful islet yield and graft function. *Diabetologia*, 48:906–912.

N.Otsu (1979). A thresholding selection method from gray-scale histogram. *IEEE Trans. on System, Man, and Cybernetics*, 9:62–66.

R.M.Rangayyan (2005). *Biomedical Image Analysis*. CRC Press.

C.Ricordi, P.E.Lacy, D.W.Scharp (1989). Automated islet isolation from human pancreas diabetes. *Diabetes*, 38(1):140–142.

C.Ricordi, R.Alejandro, H.H.Rilo, P.B.Carroll, A.G.Tzakis, T.E.Starzl, D.H.Mintz (1995). Long-term in vivo function of human mantled islets obtained by incomplete pancreatic dissociation and purification. *Transplant Proc.*, 27(6):3382.

C.Ricordi, J.R.Lakey, B.J.Hering (2001). Challenges toward standardization of islet isolation technology. *Transplant Proc.*, 33(1–2):1709.

E.A.Ryan, B.W.Paty, P.A.Senior, D.Bigam, E.Alfadhli, N.M.Kneteman, J.R.Lakey, A.M.Shapiro (2005). Five-year follow-up after clinical islet transplantation. *Diabetes*, 54(7):2060–2069.

A.M.Shapiro, C.Ricordi, B.J.Hering (2006). International trial of the edmonton protocol for islet transplantation. *The New England Journal of Medicine*, 355(13):1318–1330.

P.Soille (2003). *Morphological Image Analysis*. Springer-Verlag, New York, 2nd edition.

C.N.Street, J.R.Lakey, A.M.Shapiro, S.Imes, R.V.Rajotte, E.A.Ryan, J.G.Lyon, T.Kin, J.Avila, T.Tsujimura, G.S.Korbutt (2004). Islet graft assessment in the edmonton protocol: implications for predicting long-term clinical outcome. *Diabetes*, 53:3107–3114.

The metamorph system, © molecular devices and universal imaging corporation, WWW.UNIVERSALIMAGING.COM.

C.Tripodo, C.Valenti, B.Ballarò, Z.Rudzki, D.Tegolo, V.Gesù, A.M.Florena, V.Franco (2006). Megakaryocytic features useful for the diagnosis of myeloproliferative disorders can be obtained by a novel unsupervised software analysis. *Histology and Histopathology - Cellular and Molecular Biology*, 21:813–821.

F.Vargas, M.Vives-Pi, N.Somoza (1996). Advantages of using a cell separator and metrizamide gradients for human islet purification. *Transplantation*, 61:1562–1566.

NOISE REDUCTION AND VOICE SEPARATION ALGORITHMS APPLIED TO WOLF POPULATION COUNTING

B. Dugnol, C. Fernández, G. Galiano and J. Velasco
Dpto. de Matemticas, Universidad de Oviedo, c/ Calvo Sotelo s/n, 33007 Oviedo, Spain
dugnol@uniovi.es, carlos@uniovi.es, galiano@uniovi.es, julian@uniovi.es

Keywords: Time-frequency distribution, instantaneous frequency, signal separation, noise reduction, chirplet transform, partial differential equation, population counting.

Abstract: We use signal and image theory based algorithms to produce estimations of the number of wolves emitting howls or barks in a given field recording as an individuals counting alternative to the traditional trace collecting methodologies. We proceed in two steps. Firstly, we clean and enhance the signal by using PDE based image processing algorithms applied to the signal spectrogram. Secondly, assuming that the wolves chorus may be modelled as an addition of nonlinear chirps, we use the quadratic energy distribution corresponding to the Chirplet Transform of the signal to produce estimates of the corresponding instantaneous frequencies, chirp-rates and amplitudes at each instant of the recording. We finally establish suitable criteria to decide how such estimates are connected in time.

1 INTRODUCTION

Wolf is a protected specie in many countries around the world. Due to their predator character and to their proximity to human settlements, wolves often kill cattle interfering in this way in farmers' economy. To smooth this interference, authorities reimburse the cost of these lost to farmers. Counting the population of wolves inhabiting a region is, therefore, not only a question of biological interest but also of economic interest, since authorities are willing to estimate the budget devoted to costs produced by wolf protection, see for instance (Skonhoft, 2006). However, estimating the population of wild species is not an easy task. In particular, for mammals, few and not very precise techniques are used, mainly based on the recuperation of field traces, such as steps, excrements and so on. Our investigation is centered in what it seems to be a new technique, based on signal and image theory methods, to estimate the population of species which fulfill two conditions: living in groups, for instance, packs of wolves, and emitting some characteristic sounds, howls and barks, for wolves. The basic initial idea is to produce, from a given recording, some time-frequency distribution which allows to identify the different howls corresponding to different individuals by estimating the instantaneous frequency (IF) lines of their howls.

Unfortunately, the real situation is somehow more involved due mainly to the following two factors. On one hand, since natural sounds, in particular wolf howling, are composed by a fundamental pitch and several harmonics, direct instantaneous frequency estimation of the multi-signal recording leads to an over-counting of individuals since various IF lines correspond to the same individual. Therefore, more sophisticated methods are indicated for the analysis of these signals, methods capable of extracting additional information such as the slope of the IF, which allows to a better identification of the harmonics of a given fundamental tone. The use of a Chirplet type transform (S. Mann, 1995; L. Angrisani, 2002) is investigated in this article, although an equivalent formulation in terms of the Fourier fractional transform (H. M. Ozaktas, 2001) could be employed as well. On the other hand, despite the quality of recording devices, field recordings are affected for a variety of undesirable signals which range from low amplitude broad spectrum long duration signals, like wind, to signals localized in time, like cattle bells, or localized

in spectrum, like car engines. Clearly, the addition of all these signals generates an unstructured noise in the background of the wolves chorus which impedes the above mentioned methods to work properly, and which should be treated in advance. We accomplish this task by using PDE-based techniques which transforms the image of the signal spectrogram into a smoothed and enhanced approximation to the reassigned spectrogram introduced in (K. Kodera, 1978; F. Auger, 1995) as a spectrogram readability improving method.

2 SIGNAL ENHANCEMENT

In previous works (B. Dugnol, 2007a; B. Dugnol, 2007d), we investigated the noise reduction and edge (IF lines) enhancement on the spectrogram image by a PDE-based image processing algorithm. For a clean signal, the method allows to produce an approximation to the reassigned spectrogram through a process referred to as *differential reassignment*, and for a noisy signal this process is modified by the introduction of a nonlinear operator which induces isotropic diffusion (noise smoothing) in regions with low gradient values, and anisotropic diffusion (edge-IF enhancement) in regions with high gradient values.

Let $x \in L^2(\mathbb{R})$ denote an audio signal and consider the Short Time Fourier transform (STFT)

$$\mathcal{G}_\varphi(x;t,\omega) = \int_{\mathbb{R}} x(s)\varphi(s-t)e^{-i\omega s}ds, \qquad (1)$$

corresponding to the real, symmetric and normalized *window* $\varphi \in L^2(\mathbb{R})$. The energy density function or *spectrogram* of x corresponding to the window φ is given by

$$S_\varphi(x;t,\omega) = |\mathcal{G}_\varphi(x;t,\omega)|^2, \qquad (2)$$

which may be expressed also as (Mallat, 1998)

$$S_\varphi(x;t,\omega) = \int_{\mathbb{R}^2} WV(\varphi;\tilde{t},\tilde{\omega})WV(x;t-\tilde{t},\omega-\tilde{\omega})d\tilde{t}d\tilde{\omega}, \qquad (3)$$

with $WV(y;\cdot,\cdot)$ denoting the Wigner-Ville distribution of $y \in L^2(\mathbb{R})$,

$$WV(y;t,\omega) = \int_{\mathbb{R}} y(t+\frac{s}{2})y(t-\frac{s}{2})e^{-iws}ds.$$

The Wigner-Ville (WV) distribution has received much attention for IF estimation due to its excellent concentration and many other desirable mathematical properties, see (Mallat, 1998). However, it is well known that it presents high amplitude sign-varying cross-terms for multi-component signals which makes its interpretation difficult. Expression (3) represents the spectrogram as the convolution of the WV distribution of the signal, x, with the smoothing kernel defined by the WV distribution of the window, φ, explaining the mechanism of attenuation of the cross-terms interferences in the spectrogram. However, an important drawback of the spectrogram with respect to the WV distribution is the broadening of the IF lines as a direct consequence of the smoothing convolution. To override this inconvenient, it was suggested in (K. Kodera, 1978) that instead of assigning the averaged energy to the geometric center of the smoothing kernel, (t,ω), as it is done for the spectrogram, one assigns it to the *center of gravity* of these energy contributions, $(\hat{t},\hat{\omega})$, which is certainly more representative of the local energy distribution of the signal. As deduced in (F. Auger, 1995), the gravity center may be computed by the following formulas

$$\hat{t}(x;t,\omega) = t - \Re\left\{\frac{\mathcal{G}_{T\varphi}(x;t,\omega)}{\mathcal{G}_\varphi(x;t,\omega)}\right\},$$
$$\hat{\omega}(x;t,\omega) = \omega + \Im\left\{\frac{\mathcal{G}_{D\varphi}(x;t,\omega)}{\mathcal{G}_\varphi(x;t,\omega)}\right\},$$

where the STFT's windows in the numerators are $T\varphi(t) = t\varphi(t)$ and $D\varphi(t) = \varphi'(t)$. The reassigned spectrogram, $RS_\varphi(x;t,\omega)$, is then defined as the aggregation of the reassigned energies to their corresponding locations in the time-frequency domain. Observe that energy is conserved through the reassignment process. Other desirable properties, among which non-negativity and perfect localization of linear chirps, are proven in (Auger, 1991). For our application, it is of special interest the fact that the reallocation vector, $\mathbf{r}(t,\omega) = (\hat{t}(t,\omega)-t, \hat{\omega}(t,\omega)-\omega)$, may be expressed through a potential related to the spectrogram (E. Chassandre-Mottin, 1997),

$$\mathbf{r}(t,\omega) = \frac{1}{2}\nabla\log(S_\varphi(x;t,\omega)), \qquad (4)$$

when φ is a Gaussian window of unit variance. Let $\tau \geq 0$ denote an artificial time and consider the dynamical expression of the reassignment given by $\Phi(t,\omega,\tau) = (t,\omega)+\tau\mathbf{r}(t,\omega)$ which, for $\tau = 0$ to $\tau = 1$, connects the initial point (t,ω) with its reassigned point $(\hat{t},\hat{\omega})$. Rewriting this expression as

$$\frac{1}{\tau}(\Phi(t,\omega,\tau)-\Phi(0,\omega,\tau)) = \mathbf{r}(t,\omega),$$

and taking the limit $\tau \to 0$, we may identify the displacement vector $\mathbf{r}$ as the velocity field of the transformation Φ. In close relation with this approach is the process referred to as *differential reassignment* (E. Chassandre-Mottin, 1997), defined as the transformation given by the dynamical system corresponding

to such velocity field,

$$\begin{cases} \dfrac{d\chi}{d\tau}(t,\omega,\tau) = \mathbf{r}(\chi(t,\omega,\tau)), \\ \chi(t,\omega,0) = (t,\omega), \end{cases} \tag{5}$$

for $\tau > 0$. Observe that, in a first order approximation, we still have that χ connects (t,ω) with some point in a neighborhood of $(\hat{t},\hat{\omega})$, since

$$\begin{aligned} \chi(t,\omega,1) &\approx \chi(t,\omega,0) + \mathbf{r}(\chi(t,\omega,0)) \\ &= (t,\omega) + \mathbf{r}(t,\omega) = (\hat{t},\hat{\omega}). \end{aligned}$$

In addition, for $\tau \to \infty$, each particle (t,ω) converges to some local extremum of the potential $\log(S_\varphi(x;\cdot,\cdot))$, among them the maxima and ridges of the original spectrogram. The conservative energy reassignation for the differential reassignment is obtained by solving the following problem for $u(t,\omega,\tau)$ and $\tau > 0$,

$$\frac{\partial u}{\partial \tau} + \operatorname{div}(u\mathbf{r}) = 0, \tag{6}$$

$$u(\cdot,\cdot,0) = u_0, \tag{7}$$

where we introduced the notation $u_0 = S_\varphi(x;\cdot,\cdot)$ and, consequently, $\mathbf{r} = \frac{1}{2}\nabla\log(u_0)$. Since in applications both signal and spectrogram are defined in bounded domains, we assume (6)-(7) to hold in a bounded time-frequency domain, Ω, in which we assume non energy flow conditions on the solution and the data

$$\nabla u \cdot \mathbf{n} = 0, \quad \mathbf{r}\cdot\mathbf{n} = 0 \quad \text{on } \partial\Omega \times \mathbb{R}_+, \tag{8}$$

being $\mathbf{n}$ the unitary outwards normal to $\partial\Omega$. Finally, observe that the positivity of the spectrogram (Mallat, 1998) and the fact that it is obtained from a convolution with a C^∞ kernel implies the regularity $u_0, \mathbf{r} \in C^\infty$ and, therefore, problem (6)-(8) admits a unique smooth solution.

As noted in (E. Chassandre-Mottin, 1997), differential reassignment can be viewed as a PDE based processing of the spectrogram image in which the energy tends to concentrate on the initial image ridges (IF lines). As mentioned above, our aim is not only to concentrate the diffused IF lines of the spectrogram but also to attenuate the noise present in our recordings. It is clear that noise may distort the reassigned spectrogram due to the change of the energy distribution and therefore of the gravity centers of each time-frequency window. Although even a worse situation may happen to the differential reassignment, due to its convergence to spectrogram local extrema (noise picks among them) an intuitive way to correct this effect comes from its image processing interpretation. As shown in (B. Dugnol, 2007a; B. Dugnol, 2007d), when a strong noise is added to a clean signal better results are obtained for approximating the clean spectrogram if we use a noise reduction edge enhancement

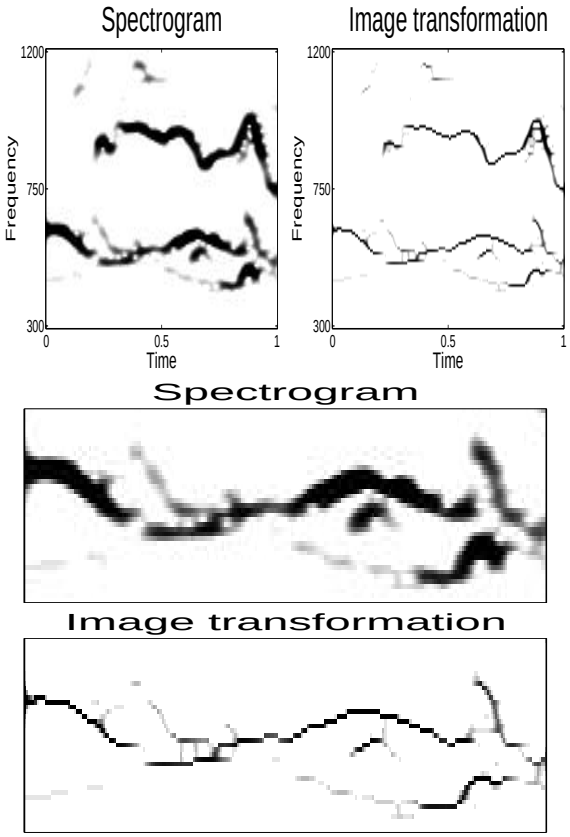


Figure 1: First row: Spectrogram and its transformation with the PDE model. Subsequent plots: detail of the howl contained within the range $300-750$ Hz. We observe the IF concentration and smoothing effect of the PDE algorithm.

PDE based algorithm than if we simply threshold the image spectrogram. This is due to the local application of gaussian filters in regions of small gradients (noise, among them) while anisotropic diffusion (in the orthogonal direction to the gradient) is applied in regions of large gradients (edges-IF lines). Therefore, a possible way to improve the image obtained by the differential reassigned spectrogram is modifying (6) by adding a diffusive term with the mentioned properties.

Let us make a final observation before writing the model we work with. In the derivation of both the reassigned and the differential reassigned spectrogram the property of energy conservation is imposed, implying that energy values on ridges increase. Indeed, let B be a neighborhood of a point of maximum for u_0, in which $\operatorname{div}\mathbf{r} = \Delta\log u_0 < 0$, and let $(t_0,\omega_0) \in B$. Let $\chi_0(t,\omega,\tau)$ denote the characteristic defined by (5) starting at (t_0,ω_0). Evaluating Eq. (6) along χ_0 we obtain

$$\frac{d}{d\tau}u = \frac{\partial u}{\partial \tau} + \mathbf{r}\cdot\nabla u = -u\operatorname{div}\mathbf{r}, \tag{9}$$

implying that u experiments exponential increase in

B. For image processing, it is desirable the maximum principle to hold, i.e., that the bounds $\min u_0 \leq u \leq \max u_0$ hold for any $(t,\omega,\tau) \in \Omega \times \mathbb{R}_+$, ensuring that the processed image lies within the range of image definition ($[0,255]$, usually). A simple way, which we shall address, to ensure this property is by dropping the right hand side term of Eq. (9), i.e., replacing Eq. (6) by the transport equation

$$\frac{\partial u}{\partial \tau} + \mathbf{r} \cdot \nabla u = 0. \tag{10}$$

However, no energy conservation law will apply anymore (note that u is constant along the characteristics). The combination of the differential reassignment problem with the edge-detection image-smoothing algorithm (L. Álvarez, 1992) is written as

$$\frac{\partial u}{\partial \tau} + \frac{\varepsilon}{2} \nabla \log(u_0) \cdot \nabla u - g(|G_s * \nabla u|) A(u) = 0, \tag{11}$$

in $\Omega \times \mathbb{R}_+$, together with the boundary data (8) and the initial condition (7). Parameter $\varepsilon \geq 0$ allows us to play with different balances between transport and diffusion effects. The diffusion operator is given by

$$A(u) = (1 - h(|\nabla u|))\Delta u + h(|\nabla u|) \sum_{j=1,\ldots,n} f_j\left(\frac{\nabla u}{|\nabla u|}\right) \frac{\partial^2 u}{\partial x_j^2}.$$

Let us briefly remind the properties and meaning of the diffusive term components in equation (11):

- Function G_s is a Gaussian of variance s. The variance is a *scale parameter* which fixes the minimal size of the details to be kept in the processed image.
- Function g is non-increasing with $g(0) = 1$ and $g(\infty) = 0$. It is a *contrast* function, which allows to decide whether a detail is sharp enough to be kept.
- The composition of G_s and g on ∇u rules the speed of diffusion in the evolution of the image, controlling the *enhancement* of the edges and the noise smoothing.
- The diffusion operator A combines isotropic and anisotropic diffusion. The first smoothes the image by local averaging while the second enforces the diffusion only on the orthogonal direction to ∇u (along the edges). More precisely, for $\theta_j = (j-1) * \pi/n$, $j = 1,\ldots,n$ we define x_j as the orthogonal to the direction θ_j, i.e., $x_j = -t \sin\theta_j + \omega \cos\theta_j$. Then, smooth non-negative functions $f_j(\cos\theta, \sin\theta)$ are designed to be *active* only when θ is close to θ_j. Therefore, the anisotropic diffusion is taken in an approximated direction to the orthogonal of ∇u. The combination of isotropic and anisotropic diffusions is controlled by function $h(s)$, which is nondecreasing with

$$h(s) = \begin{cases} 0 & \text{for } s \leq h_0, \\ 1 & \text{for } s \geq 2h_0, \end{cases} \tag{12}$$

being h_0 the *enhancement* parameter.

In Fig. 1 we show an example of the outcome of our algorithm for a signal composed by two howls. See (B. Dugnol, 2007d; B. Dugnol, 2007c) for more details and other numerical experiments.

3 HOWL TRACKING AND SEPARATION

A wolves chorus is composed, mainly, by howls and barks which, from the analytical point of view, may be regarded as chirp functions. The former has a long time support and a small frequency range variation, while the latter is almost punctually localized in time but posses a large frequency spectrum. It is convenient, therefore, adopting a parametric model to represent the wolves chorus as an addition of chirps given by the function $f : [0,T] \to \mathbb{C}$,

$$f(t) = \sum_{n=1}^{N} a_n(t) \exp[i\phi_n(t)], \tag{13}$$

with T the length of the chorus emission, a_n and ϕ_n the chirps amplitude and phase, respectively, and with N, the number of chirps contained in the chorus. We notice that N is not necessarily the number of wolves since, for instance, harmonics of a given fundamental tone are counted separately.

To identify the unknowns N, a_n and ϕ_n we proceed in two steps. Firstly, for a time discretization of the time interval $[0,T]$, say t_j, for $j = 0,\ldots,J$, we produce estimates of the amplitude $a_n(t_j)$ and the phase $\phi_n(t_j)$ of the chirps contained at such discrete times. Secondly, we establish criteria which allow us deciding if the computed estimates at adjacent times do belong to the same global chirp or do not.

For the first step we use the Chirplet transform defined by

$$\Psi f(t_o,\xi,\mu;\lambda) = \int_{-\infty}^{\infty} f(t)\, \overline{\psi_{t_o,\xi,\mu,\lambda}(t)}\, dt, \tag{14}$$

with the complex window $\psi_{t_o,\xi,\mu,\lambda}$ given by

$$\psi_{t_o,\xi,\mu,\lambda}(t) = v_\lambda(t - t_o) \exp\left[i(\xi t + \frac{\mu}{2}(t-t_o)^2)\right]. \tag{15}$$

Here, $v \in L^2(\mathbb{R})$ denotes a real window, $v_\lambda(\cdot) = v(\cdot/\lambda)$, with $\lambda > 0$, and the parameters $t_o, \xi, \mu \in \mathbb{R}$, stand for time, instantaneous frequency and chirp rate, respectively. The quadratic energy distribution corresponding to the chirplet transform (14) is given by

$$P_\Psi f(t_o,\xi,\mu;\lambda) = |\Psi f(t_o,\xi,\mu;\lambda)|^2. \tag{16}$$

For a linear chirp of the form

$$f(t) = a(t)\exp[i(\frac{\alpha}{2}(t-t_o)^2 + \beta(t-t_o) + \gamma)],$$

it is straightforward to prove that the energy distribution (16) has a global maximum at (α,β), allowing us to determine the IF and chirp rate of a given linear chirp by localizing the maxima of the energy distribution. For more general forms of mono-component chirps we have the following localization result (B. Dugnol, 2007b)

Theorem 1 *Let* $f(t) = a(t)\exp[i\phi(t)]$, *with* $a \in L^2(\mathbb{R})$ *non-negative and* $\phi \in C^3(\mathbb{R})$. *For all* $\varepsilon > 0$ *and* $\xi, \mu \in \mathbb{R}$ *there exists* $L > 0$ *such that if* $\lambda < L$ *then*

$$P_\Psi f(t_o, \xi, \mu; \lambda) \leqslant \varepsilon + P_\Psi f\left(t_o, \phi'(t_o), \phi''(t_o); \lambda\right). \tag{17}$$

In addition,

$$\lim_{\lambda \to 0} P_\Psi f\left(t_o, \phi'(t_o), \phi''(t_o); \lambda\right) = a(t_o)^2. \tag{18}$$

In other words, for a general mono-component chirp the energy distribution maximum provides an arbitrarily close approximation to the IF and chirp rate of the signal. Moreover, its amplitude may also be estimated by shrinking the window time support at the maximum point.

Finally, for a multi-component chirp $f(t) = \sum_{n=1}^{N} a_n(t)\exp[i\phi_n(t)]$ the situation is somehow more involved since although the energy distribution still has maxima at $(\phi_n'(t_o), \phi_n''(t_o))$ for all n such that $a_n(t_o) \neq 0$, these are now of local nature, and in fact, spurious local maxima not corresponding to any chirp may appear due to the energy interaction among the actual chirps.

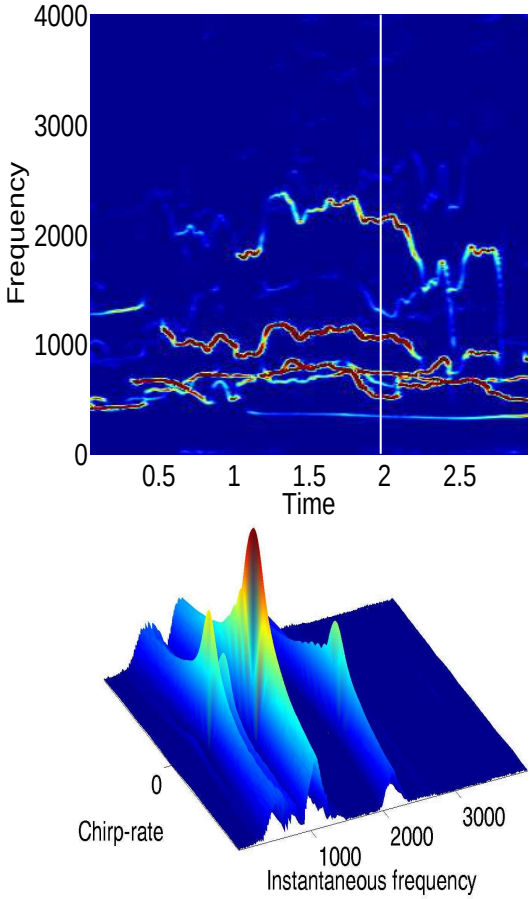


Figure 2: Above: STFT of a field recorded signal. Below: quadratic energy distribution of the chirplet transform at $t_0 = 2$. Maxima correspond to IF and chirp rate chirps locations. We observe the different behavior in the ξ and μ directions at these maxima.

4 NUMERICAL EXPERIMENTS

According to the recording quality, we start our algorithm enhancing the signal with the PDE algorithm explained in Section 2 or directly with the separation algorithm introduced in Section 3. For details about the implementation of the former, we refer the reader to (B. Dugnol, 2007c). Following, we briefly comment about the separation algorithm implementation. We start by computing the energy distribution, $P_\Psi f(\tau_m, \xi, \mu; \lambda)$ at a set of discrete times $\tau_m = m * \tau$ of constant time step, τ, and for a fixed window width λ. Next we compute the maxima of the energy at each of these times. When the signal is mono-component or the various components of the signal are far from each other relative to the window width, the maxima of P_Ψ correspond to some $(\phi_n'(\tau_m), \phi_n''(\tau_m))$ which are then identified as the IF and chirp-rate of a chirp candidate. However, when multi-component signals are close to each other or are crossing, some spurious local maxima are produced which do not correspond to any actual chirp. Therefore, some criterium must be used to select the correct local maxima at each τ_m. Although we lack of an analytical proof, there are evidences suggesting that maxima produced by chirps, i.e., at points of the type $(\phi_n'(\tau_m), \phi_n''(\tau_m))$, decrease much faster in the ξ direction than in the μ direction, see Fig. 2, a phenomenon that does not occur at spurious maxima. We use this fact to choose the candidates first by selecting ξ_k, for $k = 1, \ldots, K$, which are maxima for

$$\sup_\mu P_\Psi f(\tau_m, \xi, \mu; \lambda),$$

and, among them, selecting the maxima with respect to μ of $P_\Psi f(\tau_m, \xi_k, \mu; \lambda)$. We finally establish a threshold parameter to filter out possible local max-

ima located at points that do not correspond to any $\phi_n'(\tau_m)$ but which are close to two of them. We set this threshold such that the existence of two consecutive maxima is avoided.

In this way we obtain, for each τ_m, a set of points (μ_{i_m}, ξ_{i_m}), for $i_m = 1, \ldots, I_m$, which correspond to the IF's and chirp-rates of chirps with time support including τ_m.

The next step is the chirp separation. We note that if the time step $\tau = \tau_{m+1} - \tau_m$ is small enough, then

$$\xi_{j_{m+1}} - \tau \frac{\mu_{j_{m+1}}}{2} \approx \xi_{i_m} + \tau \frac{\mu_{i_m}}{2}.$$

Introducing a new parameter, ν, we test this property by imposing the condition

$$\frac{1}{\nu} < \frac{2\xi_{j_{m+1}} - \tau\mu_{j_{m+1}}}{2\xi_{i_m} + \tau\mu_{i_m}} < \nu, \qquad (19)$$

for two points to be in the same chirp. In the experiments we take $\nu = 2^{1/13} \approx 1.0548$.

Finally, in the case in which test (19) is satisfied by more than one point, i.e., when there exist points $\left(\xi_{j_{m+1}}, \mu_{j_{m+1}}\right)$ and $\left(\xi_{k_{m+1}}, \mu_{k_{m+1}}\right)$ such that (19) holds for the same (ξ_{i_m}, μ_{i_m}), we impose a regularity criterium and choose the point with a closer chirp-rate to that of (ξ_{i_m}, μ_{i_m}). This is a situation typically arising at chirps crossings points.

Summarizing, the chirp separation algorithm is implemented as follows:

- Each point (ξ_{i_1}, μ_{i_1}), for $i_1 = 1, \ldots, I_1$, is assumed to belong to a distinct chirp.
- For $k = 2, 3, \ldots$, we use the described criteria to decide if $\left(\xi_{i_k}, \mu_{i_k}\right)$, for : $i_k = 1, \ldots, I_k$, belongs to an already detected chirp. On the contrary, it is established as the starting point of a new chirp.
- When the above iteration is finished and to avoid artifacts due to numerical errors, we disregard chirps composed by a unique point.

Finally, once the chirps are separated, we use the following approximation, motivated by Theorem 1, to estimate the amplitude

$$a(\tau_m)^2 \approx \frac{1}{\lambda\left[\widehat{v}(0)\right]^2} P_\Psi f\left(t_o, \phi'(\tau_m), \phi''(\tau_m); \lambda\right).$$

Again, to avoid artifacts due to numerical discretization, we neglect portions of signals with an amplitude lower than certain relative threshold , $\varepsilon \in (0,1)$, of the maximum amplitude of the whole signal, considering that in this case no chirp is present.

4.1 Experiment 1. A Synthetic Signal

In this first experiment we test our algorithm with a synthetic signal, f, composed by the addition of three

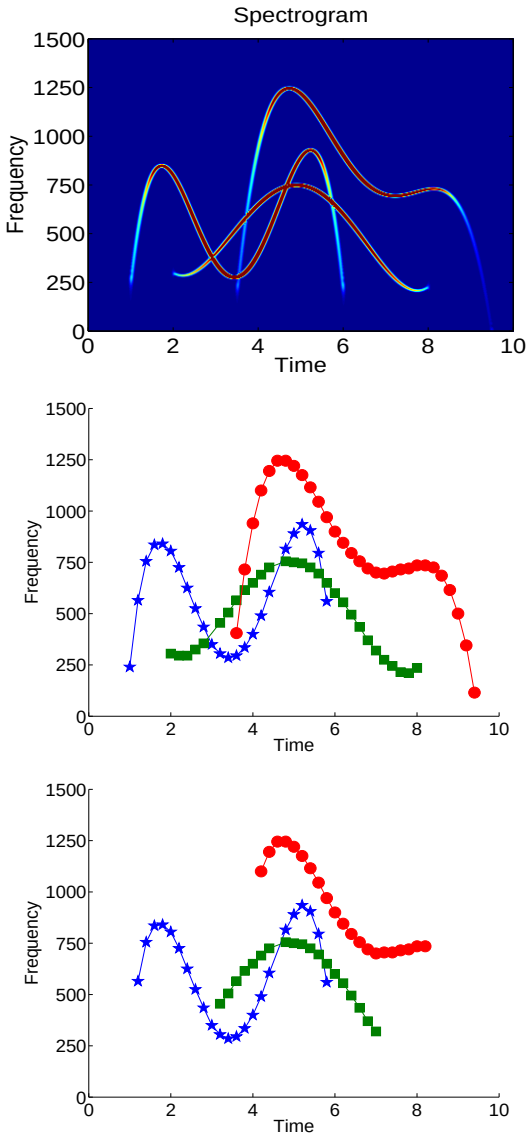


Figure 3: Spectrogram of the clean signal and results of the chirp localization and separation algorithm for clean and noisy signals, respectively.

nonlinear chirps (spectrogram shown in Fig. 3) and with the same signal corrupted with an additive noise of similar amplitude than that of f, i.e., with $SNR = 0$.

We used the same time step, $\tau = 0.2$ sec, and window width, $\lambda = 0.1$ sec, to process both signals, while we set the relative threshold amplitude level to $\varepsilon = 0.01$ for the clean signal and to $\varepsilon = 0.1$ for the noisy signal. The results of our algorithm of denoising, detection and separation is shown in Fig. 3. We observe that for the clean signal all chirps are captured with a high degree of accuracy even at crossing points. We also observe that the main effect of noise corruption is the lose of information at chirps

low amplitude range. However, the number of them is correctly computed.

In Fig. 4 we show the amplitude, IF and chirp-rate estimations of the chirp which is more affected by the noise corruption, for both clean and noisy signals. The main effect of noise corruption is observed in the amplitude computation and in the lose of information in the tails of the three quantities.

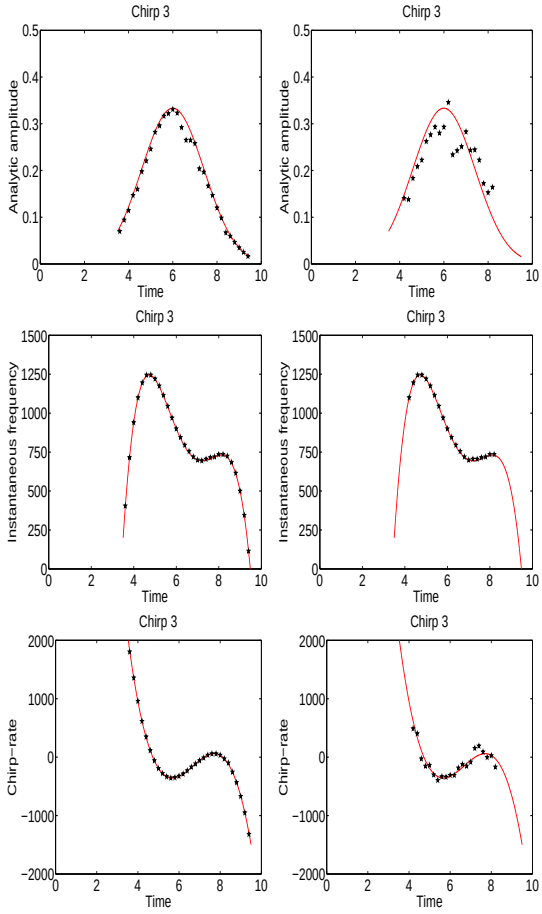


Figure 4: Amplitude, IF and chirp rate for the clean signal (left column) and noisy signal (right column). Solid lines correspond to exact values and crosses to computed values.

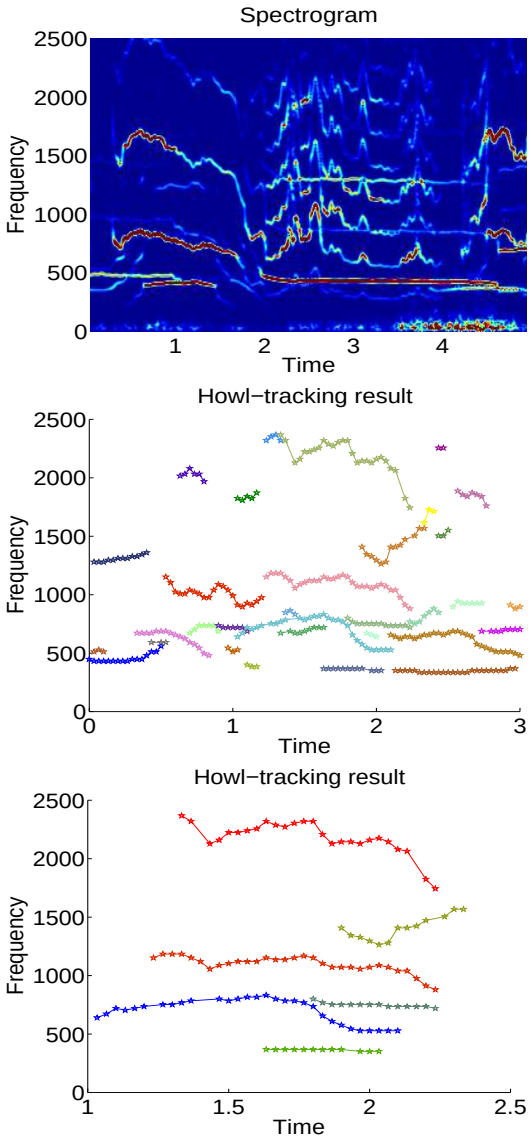


Figure 5: First row: spectrogram of the field recorded signal utilized in Experiment 2. Second row: result of the chirp localization and separation algorithm. Third row: a zoom of the previous plot showing six separated chirps corresponding to five wolves howls.

4.2 Experiment 2. A Field Recorded Wolves Chorus

In this experiment we analyze a rather complex signal obtained from field recordings of wolves choruses in wilderness, (L. Llaneza,). Due to the noise present in the recording, we first use the PDE algorithm to enhance the signal and reduce the noise, see (B. Dugnol, 2007c) for details. For the separation algorithm, we fixed the time step as $\tau = 0.03$ sec, the relative am-

plitude threshold as $\varepsilon = 0.01$ and the window width as $\lambda = 0.0625$ sec.

The algorithm output is composed by 32 chirps which should correspond to the howls and barks (with all their harmonics) emitted by the wolves along the duration of the recording (about five sec). The result is shown in Fig. 5. Since our aim is giving an estimate of how many individuals are emitting in a recording, we plot a zoom of the separating algorithm result for the time interval $(1, 2.5)$. Here, the number of chirps reduces to six. However, it seems that one couple of them are harmonics, the couples formed by the chirp around 1000 Hz and the highest IF chirp. Therefore, we may conclude that at least five wolves are emitting in this interval of time. A similar analysis is carried out with other time subintervals until all the recorded signal is analyzed.

5 CONCLUSIONS

A combined algorithm for signal enhancement and voice separation is utilized for wolf population counting. Although field recorded wolf chorus signals posses a complex structure due to noise corruption and nonlinear multi-component character, the outcome of our algorithm provides us with accurate estimates of the number of individuals emitting in a given recording. Thus, the algorithm seems to be a good complement or, even, an alternative to existent methodologies, mainly based in wolf traces collection or in the intrusive attaching of electronic devices to the animals. Clearly, our algorithm is not limited to wolves emissions but to any signal which may reasonably be modelled as an addition of chirps, opening its utilization to other applications. Drawbacks of the algorithm are related to the expert dependent election of some parameters, such as the amplitude threshold, or to the execution time when denoising and separation of long duration signals must be accomplished. We are currently working in the improvement of these aspects as well as in the recognition of components corresponding to the same emissor, such as harmonics of a fundamental chirp, pursuing the full automatization of the counting algorithm.

ACKNOWLEDGEMENTS

The first three authors are supported by Project PC0448, Gobierno del Principado de Asturias, Spain. Third and fourth authors are supported by the Spanish DGI Project MTM2004-05417.

REFERENCES

Auger, F. (1991). *Représentation temps-fréquence des signaux non-stationnaires: Syntheèse et contributions*. thèse de doctorat, Ecole Centrale de Nantes.

B. Dugnol, C. Fernández, G. G. (2007a). Wolves counting by spectrogram image processing. *Appl. Math. Comput.*, 186:820–830.

B. Dugnol, C. Fernández, G. G. J. V. (2007b). Implementation of a chirplet transform method for separating and counting wolf howls. Preprint 1, Dpt. Mathematics, Univ. of Oviedo, Spain.

B. Dugnol, C. Fernández, G. G. J. V. (2007c). Implementation of a diffusive differential reassignment method for signal enhancement. an application to wolf population counting. *Appl. Math. Comput.* To appear. E-version: doi:10.1016/j.amc.2007.03.086.

B. Dugnol, C. Fernández, G. G. J. V. (2007d). On pde-based spectrogram image restoration. application to wolf chorus noise reduction and comparison with other algorithms. In E. Damiani, A. Dipanda, K. Y. L. L. P. S., editor, *Signal processing for image enhancement and multimedia processing*, volume 34 of *Multimedia systems and applications*. Springer Verlag. To appear. (e-version in http://www.u-bourgogne.fr/SITIS/06/Proceedings/index.htm).

E. Chassandre-Mottin, I. Daubechies, F. A. P. F. (1997). Differential reassignment. *IEEE Signal Processing Letters*, 4(10):293–294.

F. Auger, P. F. (1995). Improving the readability of time-frequency and time-scale representations by the method of reassignment. *IEEE Trans. Signal Processing*, 43(5):1068–1089.

H. M. Ozaktas, Z. Zalevsky, M. A. K. (2001). *The Fractional Fourier Transform with Applications in Optics and Signal Processing*. Wiley, Chichester.

K. Kodera, R. Gendrin, C. d. V. (1978). Analysis of time-varying signals with small bt values. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):64–76.

L. Álvarez, P. L. Lions, J. M. M. (1992). Image selective smoothing and edge detection by nonlinear diffusion. ii. *SIAM J. Numer. Anal.*, 29(3):845–866.

L. Angrisani, M. D. (2002). A measurement method based on a modified version of the chirplet transform for instantaneous frequency estimation. *IEEE Trans. Instrum. Meas.*, 51:704–711.

L. Llaneza, V. P. Field recordings obtained in wilderness in asturias (spain) in the 2003 campaign. Asesores en Recursos Naturales, S.L.

Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press, London.

S. Mann, S. H. (1995). The chirplet transform: Physical considerations. *IEEE Trans. Signal Processing*, 43(11):2745–2761.

Skonhoft, A. (2006). The costs and benefits of animal predation: An analysis of scandinavian wolf recolonization. *Ecological Economics*, 58(4):830–841.

USE OF CEPSTRUM-BASED PARAMETERS FOR AUTOMATIC PATHOLOGY DETECTION ON SPEECH

Analysis of Performance and Theoretical Justification

Rubén Fraile, Juan Ignacio Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Osma-Ruiz
Department of Circuits & Systems Engineering, Universidad Politécnica de Madrid
Carretera de Valencia Km 7, 28031 Madrid, Spain
rfraile@ics.upm.es, igodino@ics.upm.es, nicolas.saenz@upm.es, vosma@ics.upm.es

Pedro Gómez-Vilda
Department of Computer Systems' Architecture and Technology, Universidad Politécnica de Madrid
Campus de Montegancedo s/n, Boadilla del Monte, 28660 Madrid, Spain
pedro@pino.datsi.fi.upm.es

Keywords: Speech analysis, Pattern classification.

Abstract: The majority of speech signal analysis procedures for automatic pathology detection mostly rely on parameters extracted from time-domain processing. Moreover, calculation of these parameters often requires prior pitch period estimation; therefore, their validity heavily depends on the robustness of pitch detection. Within this paper, an alternative approach based on cepstral-domain processing is presented which has the advantage of not requiring pitch estimation, thus providing a gain in both simplicity and robustness. While the proposed scheme is similar to solutions based on Mel-frequency cepstral parameters, already present in literature, it has an easier physical interpretation while achieving similar performance standards.

1 INTRODUCTION

Analysis of recorded speech is an attractive method for pathology detection since it is a low-cost non-invasive diagnostic procedure (Boyanov and Hadjitodorov, 1997). Although there is a wide range of causes for pathological voice (functional, neural, laryngeal, etc.) and a correspondingly wide range of acoustic parameters has been proposed for its detection (see (Jackson-Menaldi, 2002) for summarising tables and typical values), these intend to detect speech signal features that may be roughly classified in only three classes (Godino-Llorente et al., 2006b):

- *Short-term frequency perturbations*: both in fundamental frequency and in formants.
- *Short-term amplitude perturbations*.
- *Noise* or, more specifically, speech-to-noise ratio.

Calculation of above-mentioned acoustic parameters requires previous and reliable detection of speech fundamental frequency (pitch) (Deliyski, 1993) (Boyanov and Hadjitodorov, 1997). Nevertheless, pitch detection is not an easy task due to its sensitiveness to noise, signal distorsion, speech formants, etc. (Boyanov et al., 1993).

An alternative approach to speech signal analysis is doing it in cepstral domain, more specifically in Mel-frequency cepstral domain. Such approach, consisting in classifying patterns of so-called Mel-frequency cepstral coefficients (MFCC), does not require prior pitch estimation and has proven to be fairly robust against different kinds of speech distortion (Bou-Ghazale and Hansen, 2000), including that of telephone channel (Fraile et al., 2007), and reasonably independent of the particular way in which computations may be implemented (Ganchev et al., 2005). For these reasons, their application to automatic voice pathology detection has been proposed during the last years (Godino-Llorente and Gómez-Vilda, 2004). Yet, to authors' knowledge, up to now no physical explanation exists on the meaning of MFCC and their relevance on pathology detection.

Within this paper, a new scheme for automatic voice pathology detection is proposed. This lies half-way between usual cepstral domain and Mel-frequency cepstral domain. Namely, it takes profit from the conceptual interpretation of cepstral processing of speech signals (Deller et al., 1993), the pattern separation capability of cepstral distances (Rabiner and Juang, 1993) and the smoother spectrum es-

timation provided by the filter banks in MFCC calculation (Rabiner and Juang, 1993). The mathematical formulation of both cepstrum and MFCC parameters is revised in section 2, while the newly proposed set of parameters is introduced in section 3. The results from the application of these features to the detection of pathologies on voices belonging to a commercial database are reported in section 4. Last, the conclusions are presented in section 5.

2 MATHEMATICAL FORMULATION

2.1 Short-time Fourier Transform

As stated in previous section, the variability of speech signal is a key feature for pathology detection. The need for detecting such variability leads to the convenience of employing short-time techniques for speech processing. For this reason, in the following lines the mathematical framework for short-time processing of speech provided in (Deller et al., 1993) is revised.

Let $x[n]$ be a speech signal composed by N samples ($n = 0 \cdots N-1$) obtained at a sampling frequency equal to f_s; then it can be segmented in frames defined by:

$$f[n;m] = x[n] \cdot w[m-n] \tag{1}$$

where $w[n]$ is the framing window:

$$w[n] = 0 \;\; if \;\; n < 0 \;\; or \;\; n \geq L \tag{2}$$

and L is the frame length. Consequently, $f[n;m]$ has non-zero values only for $n \in [m-L+1, m]$. If consecutive speech frames are overlapped a number of l_0 samples, then m may have the following values:

$$m = L + p \cdot (L - l_0) - 1 \tag{3}$$

where p is the frame index and it is an integer such that:

$$0 \leq p \leq \frac{N-L}{L-l_0} \tag{4}$$

Considering the relation between the frame shift m and the frame index p, frames without time shift reference may be renamed as:

$$\begin{aligned} g_p[n] &= f[n+m-L+1;m] = \\ &= f[n+p\cdot(L-l_0);m] = \\ &= x[n+p\cdot(L-l_0)] \cdot w[(L-1)-n] \end{aligned} \tag{5}$$

where $n = 0 \cdots L-1$. From these speech frames, the short-term Discrete Fourier Transform (stDFT) is computed as:

$$S_p(k) = \sum_{n=0}^{N_{DFT}-1} \widetilde{g}_p[n] \cdot e^{-j \cdot \frac{2\pi}{N_{DFT}} \cdot kn} \tag{6}$$

where N_{DFT} is the number of points of the stDFT, $k = 0 \cdots N_{DFT}-1$ and:

$$\widetilde{g}_p[n] = \begin{cases} g_p[n] & if \quad 0 \leq n < L \\ 0 & otherwise \end{cases} \tag{7}$$

thus, if $N_{DFT} \geq L$ then (6) is equal to:

$$S_p(k) = \sum_{n=0}^{L-1} g_p[n] \cdot e^{-j \cdot \frac{2\pi}{N_{DFT}} \cdot kn} \tag{8}$$

The frequency values that correspond to each stDFT coefficient are:

$$f_k = \begin{cases} f_s \cdot \frac{k}{N_{DFT}} & if \quad k \leq \frac{N_{DFT}}{2} \\ f_s \cdot \frac{k-N_{DFT}}{N_{DFT}} & if \quad k > \frac{N_{DFT}}{2} \end{cases} \tag{9}$$

2.2 Short-time Cepstrum

In (Deller et al., 1993), an algorithm for computing the short-time cepstrum from the stDFT is given, under the assumption that $N_{DFT} >> L$:

$$c_p[q] = \frac{1}{N_{DFT}} \cdot \sum_{k=0}^{N_{DFT}-1} \log|S_p(k)| \cdot e^{j \cdot \frac{2\pi k}{N_{DFT}} \cdot q} \tag{10}$$

A physical interpretation of cepstrum can be derived from the discrete-time model for speech production that can also be found in (Deller et al., 1993). This model may be written in frequency domain as:

$$S\left(e^{j\Omega}\right) = E\left(e^{j\Omega}\right) \cdot G\left(e^{j\Omega}\right) \cdot H\left(e^{j\Omega}\right) \tag{11}$$

where $S\left(e^{j\Omega}\right)$ is the speech, $E\left(e^{j\Omega}\right)$ is the impulse train corresponding to the fundamental frequency and its harmonics, $G\left(e^{j\Omega}\right)$ is the glottal pulse waveform that modulates the impulse train and $H\left(e^{j\Omega}\right)$ is, herein, the combined effect of vocal tract and lip radiation. These components can be appreciated in figure 1, which corresponds to the average modulus of the short-term DFT calculated from one of the voice records belonging to the database referred in section 4.1.

The quick impulse-like variations in figure 1 correspond to the pitch harmonics $E\left(e^{j\Omega}\right)$, and the evolution of the impulse amplitude envelope is related to the glottal waveform $G\left(e^{j\Omega}\right)$ and the formants induced by the vocal tract $H\left(e^{j\Omega}\right)$. These formants correspond to the three envelope peaks with a decreasing level of energy that are centered at 750 Hz, 1375 Hz and 3000 Hz. In fact, these center frequencies

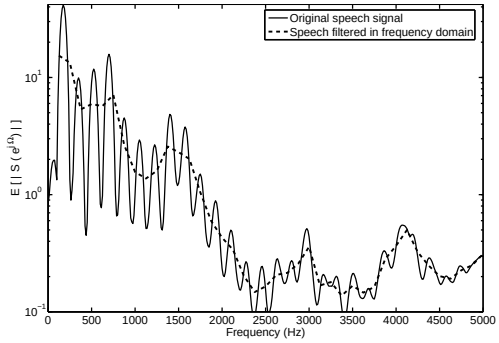


Figure 1: Average modulus of the short-term DFT for one voice record.

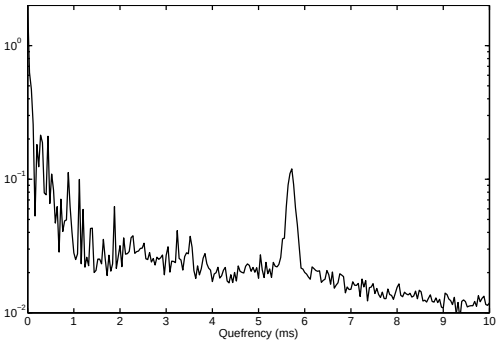


Figure 2: Short term cepstrum averaged for all frames of the same voice record as used for figure 1.

are coherent with the range of typical values given in (Jackson-Menaldi, 2002).

The logarithm operation in (10) converts the products in (11) into sums. Consequently, it allows the cepstrum to separate fast from slow signal variations in frequency domain. This widely known fact is illustrated in figure 2, where the peak around 5.7 ms clearly identifies the fundamental frequency (175 Hz) and the values below 2 ms correspond to the spectrum envelope.

2.3 Short-time MFCC

Once the stDFT of a speech signal is available, another option for further processing, as mentioned in section 1, is the calculation of short-time MFCC (stMFCC) parameters. For stMFCC computation, only the positive part of the frequency axis is considered (Rabiner and Juang, 1993), that is, $f_k \geq 0$ and, therefore, $k \leq N_{DFT}/2$. In order to calculate stMFCC coefficients, a transformation is applied to the frequencies so as to convert them to Mel-frequencies f_k^m (Godino-Llorente and Gómez-Vilda, 2004):

$$f_k^m = 2595 \cdot \log_{10}\left(1 + \frac{f_k}{700}\right) \qquad (12)$$

and the stDFT is further processed through band-pass integration along M equally long Mel-frequency intervals, being $M = \lfloor 3 \cdot \log_{10} f_s \rfloor$ ($\lfloor \cdot \rfloor$ means rounding to the previous integer). Namely, the i^{th} interval ($i = 1 \cdots M$) in Mel-domain is defined by:

$$I_i^m = \left[F^m \cdot \frac{i-1}{M+1}, F^m \cdot \frac{i+1}{M+1}\right] \qquad (13)$$

where F^m is the maximum Mel-frecuency:

$$F^m = \max_k f_k^m = 2595 \cdot \log_{10}\left(1 + \frac{f_s/2}{700}\right) \qquad (14)$$

and the interval length in Mel-domain is given by:

$$L(I_i^m) = \frac{2}{M+1} \cdot F^m \qquad (15)$$

According to previous equations, the N_{DFT} stDFT coefficients are transformed to M frequency components as follows:

$$\widetilde{S}_p(i) = \sum_{f_k \in I_i} \left(1 - \frac{\left| f_k^m - F^m \cdot \frac{i}{M+1} \right|}{L(I_i^m)/2}\right) \cdot |S_p(k)| \qquad (16)$$

Last, the q^{th} ($q = 1 \cdots Q$) stMFCC of the p^{th} speech frame, where Q is the desired length of the Mel-cepstrum, is given by cosine transform of the logarithm of the smoothed "Mel-spectrum" (Rabiner and Juang, 1993):

$$\widetilde{c}_p[q] = \sum_{i=1}^{M} \log\left|\widetilde{S}_p(i)\right| \cdot \cos\left[q \cdot \left(i - \frac{1}{2}\right) \cdot \frac{\pi}{M}\right] \qquad (17)$$

3 CEPSTRAL COEFFICIENTS BASED ON SMOOTHED SPECTRUM

3.1 Justification

As stated in section 1, while MFCC parameters exhibit both good performance and robustness in feature extraction from speech, they lack a clear physical interpretation. On the opposite, cepstrum has a physical meaning (recall section 2.2), yet raw cepstrum coefficients are not as useful for speech parametrisation. In the next paragraphs, the reasons for these facts are exposed.

Cepstrum calculation, as formulated in (10), is based on the spectrum estimate provided by the absolute value of the stDFT. Due to the logarithm, this gives a result that is proportional to the case of periodogram-based spectrum estimation. However, such estimation is very dependent on the specific values of the original speech frame. A more robust spectrum estimate can be obtained by smoothing of the

periodogram (Blackman and Tukey method, (Proakis and Manolakis, 1996)). In fact, this is what (16) expresses in the calculation of MFCC. Therefore, filtering of the stDFT may be assumed to be one of the sources of MFCC robustness.

In contrast, an explanation for the lack of clear interpretation of MFCC also lies in the meaning of (16). According to that equation, stDFT smoothing for MFCC computation is carried out with a variable-length filter, that is, a Bartlett window whose length decreases for lower frequency bands. Moreover, the smoothed stDFT is downsampled to obtain only M samples in the interval $[0, f_s/2]$ that are not uniformly spaced (Rabiner and Juang, 1993). While the downsampling is positive in the sense that it reduces the dimensionality of the problem, its non-uniformness, together with the previous variable-length filtering, obscures the interpretation of the output of the cosine transform in (17).

From the previous reasoning, if stDFT is smoothed with a fixed-length filter and its output is uniformly decimated prior to the logarithm computation, the cepstral coefficients in (10) can be transformed to a more robust parameter set. Moreover, this is achieved while keeping the physical meaning of cepstrum, since the output of the first operation gives an improved spectrum estimate and the second only limits the length of cepstrum in quefrency domain.

3.2 Formulation

Starting from (8), if the stDFT modulus is smoothed with a Bartlett window of constant length equal to Δf then the following output is obtained:

$$S'_p(i) = \sum_{f_k \in I_i} \left(1 - \frac{\left|f_k^m - i \cdot \Delta f/2\right|}{\Delta f/2}\right) \cdot |S_p(k)| \quad (18)$$

where $I_i = [\Delta f \cdot (i-1)/2, \Delta f \cdot (i+1)/2]$ and the Bartlett window has been chosen for similarity with (16). Herein, only the positive part of the frequency axis has been considered, as in section 2.3.

If the filtered stDFT is decimated so as to keep only the outputs of consecutive windows with a 50% overlap, this is equivalent to decimation by a factor $D = \lfloor \Delta f \cdot N_{DFT}/(2 \cdot f_s) \rfloor$. The modified cepstrum then becomes:

$$c'_p[q] = \frac{D}{N_{DFT}} \cdot \sum_{k=0}^{\frac{N_{DFT}}{2 \cdot D}} \log\left|S'_p(k \cdot D)\right| \cdot \cos\left((k-1) \cdot \frac{2\pi D}{N_{DFT}} \cdot q\right) \quad (19)$$

where only the positive frequencies have been considered, hence computing the inverse DFT as a cosine transform as in (17). $c'_p[q]$ has the twofold advantage over $c_p[q]$ of being based on a smoother spectrum estimate $S'_p(i)$ and having a period length that has been reduced by a factor D, thus providing some dimensionality reduction.

3.3 Cepstral Distances

Differences in cepstrum can be used for speech signal classification. An example of such usage is the definition of the cepstral distance in (Rabiner and Juang, 1993) as the norm of the vector resulting form substraction of the two cepstra to be compared. This, if directly applied to pathology detection, would result in comparing the cepstrum of consecutive speech frames so as to assess the variability of the signal. Mathematically:

$$d_p^2 = \sum_{q=0}^{\frac{N_{DFT}}{D}-1} \left|c'_{p+1}[q] - c'_p[q]\right|^2 \quad (20)$$

However, bearing in mind the physical interpretation of cepstrum, this definition has the drawback of mixing pitch variations with formant and glottal pulse variations. To overcome this problem an individual frame-to-frame cepstral parameter variation analysis is proposed:

$$d_p[q] = \left|c'_{p+1}[q] - c'_p[q]\right| \quad (21)$$

This way, analysis of the distribution of $d_p[q]$ related to speech formant and glottal pulse variability (low values of q) can be isolated from pitch changes associated to values of q around the pitch period.

4 APPLICATION AND RESULTS

For the purpose of performance analysis, the modified cepstral parameters presented in previous section have been applied to the problem of automatic pathology detection on recorded voice. The results have been compared to those produced by MFCC. Within this section, first the voice database is presented, second the used parameter set is specified, third the classifier is described and, last, the results are shown and commented.

4.1 Database

The voice records used in this investigation are the same as in (Godino-Llorente et al., 2006a). They be-

long to a database distributed by the company Kay Elemetrics (Kay Elemetrics Corp., 1994). The recorded sounds correspond to sustained phonations (1-3 s long) of the vowel /ah/ from patients with either normal or disordered voice. Such voice disorders belong to a wide variety of organic, neurological, traumatic and psychogenic classes. Sampling rate of speech records has been made uniform for all of them and equal to 25 kHz, while the coding has a resolution of 16 bits. The subset taken from the database contains 53 normal and 173 pathological speakers which are uniformly distributed in age and gender (Godino-Llorente et al., 2006a).

4.2 Parameter Sets

For each speech record, cepstrum-based coefficients, as defined in (19), have been calculated. Namely, a filter length $\Delta f = 200\ Hz$ has been chosen for sfDFT smoothing. As a consequence, a cepstrum length of $(f_s - \Delta f/2)/(\Delta f/2) = 124$ samples results. The choice of Δf is consistent to the approximate length of the low-band filters used for MFCC calculation (recall (16)). At first sight, however, it has the drawback of loosing pitch information of the signal spectrum. This is illustrated in figure 1 where the filtered DFT has been plotted with a dashed line. Nevertheless, such filtered spectrum contains information on both harmonic-to-noise ratio (HNR) and glottal pulse waveform (Murphy and Akande, 2005) and HNR is a useful parameter for pathology detection that is closely related to both frequency and amplitude perturbations of pitch (Jackson-Menaldi, 2002).

Since cepstrum contains information on total signal energy and its distribution among formants, the whole sequence is used as part of the parameter set. As well as the cepstrum, information on its variability is used as an input for the pathology detector. More specifically, the mean and variance of $d_p[q]$ for each value of q are used as descriptors of the cepstrum variability. Therefore, on the whole, a parameter vector of 124×3 elements is produced.

For the sake of comparison, another classifier based on a parameter vector consisting of $M = \lfloor 3 \cdot \log_{10} f_s \rfloor = 13$ MFCC coeficients averaged for all signal frames has also been tested.

4.3 Classifier Description

For both classification schemes, a Multilayer Perceptron (MLP) with two hidden layers, each consisting of 4 neurons, and a single-neuron output layer has been used as a classifier. All neurons have logistic activation functions. An MLP with a single hidden layer having 50 neurons was utilised in (Godino-Llorente and Gómez-Vilda, 2004). The structure herein proposed, in contrast, has less free parameters, thus allowing a faster learning, and the reduced number of neurons is compensanted by the introduction of an additional hidden layer that permits learning of more complex relations (Haykin, 1994).

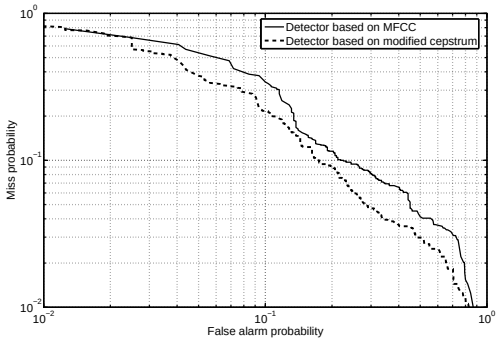


Figure 3: DET plot for MFCC based and modified cepstrum based classifiers.

4.4 Results

The MLP classifier has been trained with 70% of available speech records in such a way that its output is expected to be “1” for pathological voices and “0” for normal voices. The remaining 30% of records have been used for testing. The experiment has been repeated 20 times, each of them with different, randomly chosen, training sets. The average results for both MFCC and herein presented cepstrum-based parameters are drawn in the DET plot (Martin et al., 1997) of figure 3.

Plotted results indicate that the performance of the classifier based on the newly proposed set of parameters is in the same order of magnitude than that of MFCC parameters. To be specific, in terms of equal error rate (EER), that is, for false alarm rate equal to miss rate, the MFCC-based classification yields an experimental error probability of 15% while the cepstrum-based classification error probability for the same conditions is 14%. Considering that within this experiment the task of fine-tuning the classifier has not been carried out and that the MLP has been chosen as a standard for comparison, the difference in the results is not significant.

In order to acquire a deeper understanding of the reasons for these results, an analysis of the relevance of cepstrum-based parameters for speech classification as either pathological or not has been realised. Such analysis is based on the evaluation of the Fisher criterion (Duda et al., 2001) for each individual parameter. The results, differentiated for the three subsets of parameters (modified cepstrum, variance of

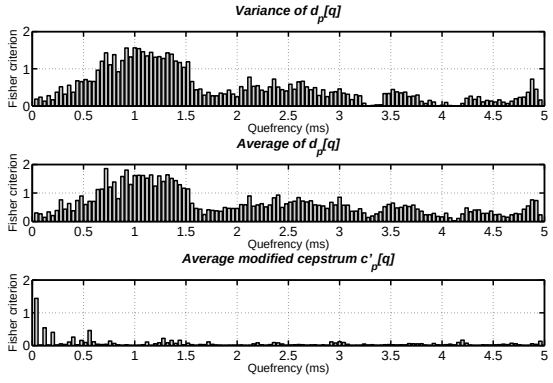


Figure 4: Value of Fisher criterion for each cepstral parameter.

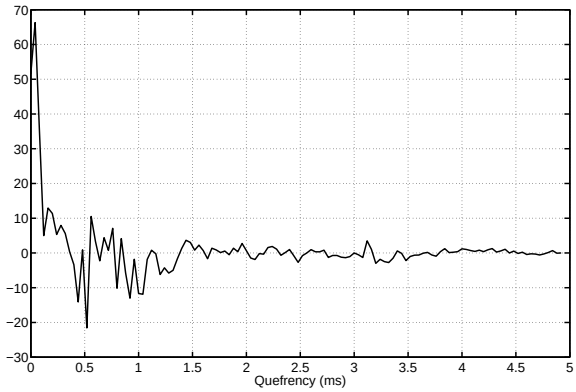


Figure 5: 124 modified cepstral parameters from one of the database's voice records.

differences and average of absolute differences) are plotted in figure 4.

According to this plot, the most relevant cepstral parameters for pathology detection maybe roughly classified into two groups:

- The modified cepstrum values with lowest indices (plot at the bottom of figure 4): these are related to the slowest components of the spectrum envelope in figure 1, which, on their side, are associated to spectral noise levels and HNR (Murphy and Akande, 2005).
- The frame-to-frame variations in cepstrum-based coefficients whose quefrecies are within the interval $[0.5, 1.5]$ miliseconds approximately: coefficients within that interval correspond to the short frequency range components of the spectrum envelope. These components, as justified in section 2.2, are related to glottal waveform and speech formants. However, this information itself does not help to discriminate the presence of pathology, as indicated by the low values of the Fisher criterion in the bottom plot of figure 4. Instead, frame-to-frame variations of these factors are much more relevant, as depicted in the other two plots of the same figure.

To be more specific, since the voice records of the database used for this experiment correspond to sustained vowel phonations, it can be assumed that the vocal tract has very little variations, hence formants do not change and the second group of parameters should be more closely related to changes in the glottal waveform. As for the limits of the quefrency interval in which parameters from the second group are relevant, the lower limit of 0.5 ms corresponds to the quefrency band that separates slow components of the spectrum envelope (first group of parameters) from faster components (associated to the second set); on the other hand, the upper limit of 1.5 ms corresponds to the highest quefrency range at which the modified cepstrum $c'_p[q]$ has significant values. This is shown in figure 5, where a plot of the frame-averaged modified cepstrum of one voice record is depicted.

5 CONCLUSIONS

Speech parametrisation in cepstral domain is a useful technique for automatic pathology detection. Specifically, MFCC have been successfully used for this purpose. While the computation of these parameters has an intrinsic robustness due to its independency from pitch extraction and the spectrum filtering, their physical interpretation is obscure because of the non-linear Mel-frequency transformation.

Within this paper an alternative set of cepstrum-based parameters has been proposed. Such parameters share the robustness of MFCC since they do not require pitch estimation and filtering of the estimated speech spectrum is also performed. In contrast to MFCC, the calculation of these newly proposed parameters does not involve any non-linear frequency transformation and, consequently, their physical interpretation remains clear. Namely, their values have been shown to be related to the amount of noise energy present in speech and the glottal waveform variability. Both factors are directly associated to laringeal pathologies.

Finally, the performance of the proposed cepstral parameters for pathology detection has been tested using a MLP classifier and results have been compared to those of MFCC. The obtained misclassification rates indicate that the performances of both sets of parameters are similar. Moreover, a deeper analysis on the individual impact of each parameter on the classification task has revealed that the most relevant parameters are those more closely linked to the abovementioned two factors: noise energy and glottal wave variations.

ACKNOWLEDGEMENTS

This research was carried out within projects funded by the Ministry of Science and Technology of Spain (TEC2006-12887-C02) and the Universidad Politécnica de Madrid (AL06-EX-PID-033).

REFERENCES

Bou-Ghazale, S. E. and Hansen, J. H. L. (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 8(4):429–442.

Boyanov, B. and Hadjitodorov, S. (1997). Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. *IEEE Engineering in Medicine and Biology*, 16(4):74–82.

Boyanov, B., Ivanov, T., Hadjitodorov, S., and Chollet, G. (1993). Robust hybrid pitch detector. *IEE Electronics Letters*, 29(22):1924–1926.

Deliyski, D. D. (1993). Acoustic model and evaluation of pathological voice production. In *Proceedings of the 3rd Conference on Speech Communication and Technology (EUROSPEECH'93)*, pages 1969–1972, Berlin (Germany).

Deller, J. R., Proakis, J. G., and Hansen, J. H. L. (1993). *Discrete-time processing of speech signals*. Macmillan Publishing Company, New York (USA).

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification*. John Wiley & sons, New York (USA), 2nd edition.

Fraile, R., Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., and Gómez-Vilda, P. (2007). Analysis of the impact of analogue telephone channel on mfcc parameters for voice pathology detection. In 8th *INTERSPEECH Conference (INTERSPEECH 2007)*, pages 1218–1221, Antwerp (Belgium).

Ganchev, T., Fakotakis, N., and Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the* 10th *International Conference on Speech and Computer (SPECOM 2005)*, pages 191–194, Patras (Greece).

Godino-Llorente, J. I. and Gómez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, 51(2):380–384.

Godino-Llorente, J. I., Gómez-Vilda, P., and Blanco-Velasco, M. (2006a). Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, 53(10):1493–1953.

Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., Aguilera-Navarro, S., and Gómez-Vilda, P. (2006b). An integrated tool for the diagnosis of voice disorders. *Medical Engineering & Physics*, 28(3):276–289.

Haykin, S. (1994). *Neural Networks: a comprehensive foundation*. Macmillan College Publishing Company, New York (USA), 1st edition.

Jackson-Menaldi, M. C. A. (2002). *La voz patológica*. Editorial Médica Panamericana, Buenos Aires (Argentina).

Kay Elemetrics Corp. (1994). Disordered voice database.version 1.03.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. (1997). The DET curve in assessment of detection task performance. In *Proceedings of the* 5th *Conference on Speech Communication and Technology (EUROSPEECH'97)*, pages 1895–1898, Rhodes (Greece).

Murphy, P. J. and Akande, O. O. (2005). Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation. In *Proceedings of the* 3th *International Conference on Non-Linear Speech Processing (NOLISP'05)*, pages 224–232, Barcelona (Spain).

Proakis, J. G. and Manolakis, D. G. (1996). *Digital Signal Processing. Principles, Algorithms and Applications*. Prentice-Hall International, New Jersey (USA), 3rd edition.

Rabiner, L. and Juang, B. H. (1993). *Fundamentals of speech recognition*. Prentice-Hall, Englewood Cliffs (USA).

FAST AND ROBUST MID-SAGITTAL PLANE LOCATION IN 3D MR IMAGES OF THE BRAIN

Felipe P. G. Bergo, Guilherme C. S. Ruppert, Luiz F. Pinto and Alexandre X. Falcão
LIV, Institute of Computing, Unicamp, C.P. 6176, Campinas, SP, 13083-970, Brazil
bergo@liv.ic.unicamp.br, ruppert@ic.unicamp.br, lfp@liv.ic.unicamp.br, afalcao@ic.unicamp.br

Keywords: Mid-sagittal plane extraction, medical image analysis, medical image alignment, brain image segmentation.

Abstract: Extraction of the mid-sagittal plane (MSP) is an important step for brain image registration and asymmetry analysis. We present a fast MSP extraction method for 3D MR images, which is based on automatic segmentation of the brain and on heuristic maximization of cerebro-spinal fluid within the MSP. The method is shown to be robust to severe anatomical asymmetries between the hemispheres, caused by surgical procedures and lesions. The experiments used 64 MR images (36 pathological, 20 healthy, 8 synthetic) and the method found an acceptable approximation of the MSP in all images with a mean time of 60.0 seconds per image.

1 INTRODUCTION

The human brain is not perfectly symmetric (Davidson and Hugdahl, 1996; Crow, 1993; Geschwind and Levitsky, 1968). However, for the purpose of analysis, it is paramount to define and distinguish a *standard of asymmetry*, considered as normal for any given measurement, from abnormal asymmetry, which may be related to neurological diseases, cerebral malformations, surgical procedures or trauma. Several works sustain this claim. For example, accentuated asymmetries between left and right hippocampi have been found in patients with Schizophrenia (Wang et al., 2001; Csernansky et al., 1998; Styner and Gerig, 2000; Mackay et al., 2003; Highley et al., 2003; Barrick et al., 2005), Epilepsy (Hogan et al., 2000; Wu et al., 2005) and Alzheimer Disease (Csernansky et al., 2000; Liu et al., 2007).

The brain can be divided in two hemispheres, and the structures of one side should have their counterpart in the other side with similar shapes and approximate locations (Davidson and Hugdahl, 1996). These hemispheres have their boundaries limited by the longitudinal (median) fissure, being the corpus callosum their only interconnection.

The ideal separation surface between the hemisferes is not perfectly planar, but the mid-sagittal plane (MSP) can be used as a reference for asymmetry analysis, without significant loss in the relative comparison between normal and abnormal subjects. The MSP location is also important for image registration. Some works have used this operation as a first step for intra-subject registration, as it reduces the number of degrees of freedom (Ardekani et al., 1997; Kapouleas et al., 1991), and to bring different images into a same coordinate system (Liu et al., 2001), such as in the Talairach (Talairach and Tournoux, 1988) model.

However, there is no exact definition of the MSP and its determination by manual delineation is sensitive to different experts. Given that, a reasonable approach for evaluation seems to be visual inspection with error quantification, when we increase the asymmetry artificially and/or linearly transform the image.

The longitudinal fissure forms a gap between the hemispheres filled with cerebro-spinal fluid (CSF). We define the MSP as a large intersection between a plane and an *envelope* of the brain (a binary volume whose surface approximates the convex hull of the brain) that maximizes the amount of CSF. This definition leads to an automatic, robust and fast algorithm for MSP extraction.

The paper is organized as follows. In Section 2, we review existing works on automatic location of the mid-sagittal plane. In section 3, we present the proposed method. In section 4, we show experimental results and validation with simulated and real MR-T1 images. Section 5 states our conclusions.

2 RELATED WORKS

MSP extraction methods can be divided in two groups: (i) methods that define the MSP as a

plane that maximizes a symmetry measure, extracted from both sides of the image (Junck et al., 1990; Minoshima et al., 1992; Sun and Sherrah, 1997; Ardekani et al., 1997; Smith and Jenkinson, 1999; Liu et al., 2001; Prima et al., 2002; Tuzikov et al., 2003; Teverovskiy and Liu, 2006), and (ii) methods that detect the longitudinal fissure to estimate the location of the MSP (Brummer, 1991; Guillemaud et al., 1996; Hu and Nowinski, 2003; Volkau et al., 2006). Table 1 summarizes these works, and extensive reviews can be found in (Hu and Nowinski, 2003), (Volkau et al., 2006), (Prima et al., 2002) and (Liu et al., 2001).

Methods in the first group address the problem by exploiting the hough symmetry of the brain. Basically, they consist in defining a symmetry measure and searching for the plane that maximizes this score. Methods in the second group find the MSP by detecting the longitudinal fissure. Even though the longitudinal fissure is not visible in some modalities, such as PET and SPECT, it clearly appears in MR images. Particularly, we prefer these methods because patients may have very asymmetric brains and we believe this would affect the symmetry measure and, consequently, the MSP detection.

The aforementioned approaches based on longitudinal fissure detection present some limitations that we are circumventing in the proposed method. In (Guillemaud et al., 1996), the MSP is found by using snakes and orthogonal regression for a set of points manually placed on each slice along the longitudinal fissure, thus requiring human intervention. Other method (Brummer, 1991) uses the Hough Transform to automatically detect straight lines on each slice (Brummer, 1991), but it does not perform well on pathological images. The method in (Hu and Nowinski, 2003) assumes local symmetry near the plane, which is not verified in many cases (see Figures 2, 5 and 8). Volkau et al. (Volkau et al., 2006) propose a method based on the Kullback and Leibler's measure for intensity histograms in consecutive candidate planes (image slices). The method presents excellent results under a few limitations related to rotation, search region of the plane, and pathological images.

3 METHODS

Our method is based on detection of the longitudinal fissure, which is clearly visible in MR images. Unlike some previous works, our approach is fully 3D, automatic, and applicable to images of patients with severe asymmetries.

We assume that the mid-sagittal plane is a plane that contains a maximal area of cerebro-spinal fluid (CSF), excluding ventricles and lesions. In MR T1 images, CSF appears as low intensity pixels, so the task is reduced to the search of a sagittal plane that minimizes the mean voxel intensity within a mask that disregards voxels from large CSF structures and voxels outside the brain.

The method is divided in two stages. First, we automatically segment the brain and morphologically remove thick CSF structures from it, obtaining a brain mask. The second stage is the location of the plane itself, searching for a plane that minimizes the mean voxel intensity within its intersection with the brain mask. Our method uses some morphological operations whose structuring elements are defined based on the image resolution. To keep the method description independent of image resolution, we use the notation S_r to denote a spherical structuring element of radius r *mm*.

3.1 Segmentation Stage

We use the tree pruning approach to segment the brain. Tree pruning (Falcão et al., 2004a; Miranda et al., 2006) is a segmentation method based on the Image Foresting Transform (Falcão et al., 2004b), which is a general tool for the design of fast image processing operators based on connectivity. In tree pruning, we interpret the image as a graph, and compute an optimum path forest from a set of seed voxels inside the object. A gradient-like image with high pixel intensities along object borders must be computed to provide the edge weights of the implicit graph. A combinatorial property of the forest is exploited to prune tree paths at the object's border, limiting the forest to the object being segmented.

To segment the brain (white matter (WM), gray matter (GM) and ventricles), we compute a suitable gradient image, a set of seed voxels inside the brain and apply the tree pruning algorithm. A more detailed description of this procedure is given in (Bergo et al., 2007). Note that any other brain segmentation method could be used for this purpose.

Gradient Computation. MR-T1 images of the brain contain two large clusters: the first with air, bone and CSF (lower intensities), and the second, with higher intensities, consists of GM, WM, skin, fat and muscles. Otsu's optimal threshold (Otsu, 1979) can separate these clusters (Figs. 1a and 1b), such that the GM/CSF border becomes part of the border between them. To enhance the GM/CSF border, we multiply each voxel intensity $I(p)$ by a weight $w(p)$ as follows:

Table 1: Summary of existing MSP methods.

Method	Based on	2D/3D	Application	Measure
(Brummer, 1991)	fissure	2D	MR	Edge Hough Transform
(Guillemaud et al., 1996)	fissure	2D	MR	Active contours
(Hu and Nowinski, 2003)	fissure	2D	MR, CT	Local symmetry of fissure
(Volkau et al., 2006)	fissure	3D	MR, CT	Kullback-Leibler's measure
(Junck et al., 1990)	symmetry	2D	PET, SPECT	Intensity cross correlation
(Minoshima et al., 1992)	symmetry	3D	PET	Stochastic sign change
(Ardekani et al., 1997)	symmetry	3D	MR, PET	Intensity cross correlation
(Sun and Sherrah, 1997)	symmetry	3D	MR, CT	Extended Gaussian image
(Smith and Jenkinson, 1999)	symmetry	3D	MR, CT, PET, SPECT	Ratio of intensity profiles
(Liu et al., 2001)	symmetry	2D	MR, CT	Edge cross correlation
(Prima et al., 2002)	symmetry	3D	MR, CT, PET, SPECT	Intensity cross correlation
(Tuzikov et al., 2003)	symmetry	3D	MR, CT, SPECT	Intensity cross correlation
(Teverovskiy and Liu, 2006)	symmetry	3D	MR	Edge cross correlation

$$w(p) = \begin{cases} 0 & I(p) \le m_1 \\ 2\left(\frac{I(p)-m_1}{m_2-m_1}\right)^2 & m_1 < I(p) \le \tau \\ 1-2\left(\frac{I(p)-m_2}{m_2-m_1}\right)^2 & \tau < I(p) \le m_2 \\ 2 & I(p) > m_2 \end{cases} \quad (1)$$

where τ is the Otsu's threshold, and m_1 and m_2 are the mean intensities of each cluster. We compute a 3D gradient at each voxel as the sum of its projections along 26 directions around the voxel, and then use its magnitude for tree pruning (Figure 1c).

Seed Selection. The brighter cluster contains many voxels outside the brain (Figure 1b). To obtain a set of seeds inside the brain, we apply a morphological erosion by S_5 on the binary image of the brighter cluster. This operation disconnects the brain from adjacent structures. We then select the largest connected component as the seed set (Figure 1d).

Morphological Closing. The brain object obtained by tree pruning (Figure 1e) might not include the entire longitudinal fissure, especially when the fissure is too thick. To ensure its inclusion, we apply a morphological closing by S_{20} to the binary brain image (Figure 1f).

Thick CSF Structure Removal. The last step of this phase is the removal of thick CSF structures (such as the ventricles, lesions and post-surgery cavities) from the brain object, to avoid the MSP from snapping to a dark structure other than the longitudinal fissure. We achieve this with a sequence of morphological operations: we start from a binary image obtained by thresholding at Otsu's optimal threshold (Figure 1b). We apply a morphological opening by

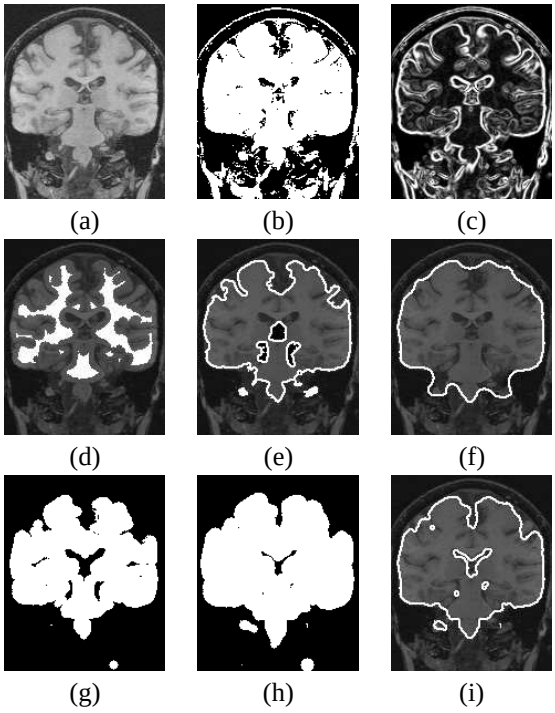


Figure 1: Sample slice of the intermediary steps in stage 1: (a) original coronal MR slice; (b) binary cluster mask obtained by thresholding; (c) gradient-like image used for tree pruning; (d) seed set used for tree pruning (white); (e) border of the brain object obtained by tree pruning (white); (f) border of the brain object after morphological closing; (g) CSF mask after opening; (h) CSF mask after dilation; (h) brain mask (intersection of (f) and (h)).

S_5 to connect the thick (> 5 *mm*) CSF structures (Figure 1g), and then dilate the result by S_2 to include a thin (2 *mm*) wall of the CSF structures (Figure 1h). This dilation ensures the reinclusion of the longitudinal fissure, in case it is removed by the opening. The binary intersection of this image with the brain ob-

ject is then used as brain mask (Figure 1i) by the next stage of our method. Only voxels within this mask are considered by stage 2. Figures 2a and 2b show how the computed brain mask excludes the large cavity in a post-surgery image, and figures 2c and 2d show how the mask excludes most of the ventricles in patients with large ventricles.

3.2 Plane Location Stage

To obtain the CSF score of a plane, we compute the mean voxel intensity in the intersection between the plane and the brain mask (Figures 3a and 3b). The lower the score, the more likely the plane is to contain more CSF than white matter and gray matter. The plane with a sufficiently large brain mask intersection and minimal score is the most likely to be the mid-sagittal plane.

To find a starting candidate plane, we compute the score of all sagittal planes in 1 *mm* intervals (which leads to 140–180 planes in usual MR datasets), and select the plane with minimum score. Planes with intersection area lower than 10 000 mm^2 are not considered to avoid selecting planes tangent to the surface of the brain. Planes with small intersection areas may lead to low scores due to alignment with sulci and also due to partial volume effect between gray matter and CSF (Figures 3c and 3d).

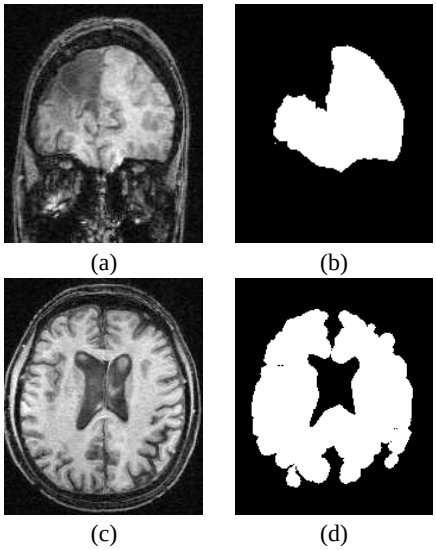


Figure 2: Examples of thick CSF structure removal: (a) coronal MR slice of a patient with post-surgical cavity; (b) brain mask of (a); (c) axial MR slice of a patient with large ventricles; (d) brain mask of (c).

Once the best candidate plane is found, we compute the CSF score for small transformations of the plane by a set of rotations and translations. If none of the transformations lead to a plane with lower CSF score, the current plane is the mid-sagittal plane and the algorithm stops. Otherwise, the transformed plane with lower CSF score is considered the current candidate, and the algorithm is repeated. The algorithm is finite, since each iteration reduces the CSF score, and the CSF score is limited by the voxel intensity domain.

We use a set of 42 candidate transforms at each iteration: translations on both directions of the X, Y and Z axes by 10 *mm*, 5 *mm* and 1 *mm* (18 translations) and rotations on both directions around the X, Y and Z axes by 10^o, 5^o, 1^o and 0.5^o (24 rotations). All rotations are about the central point of the initial candidate plane. There is no point in attempting rotations by less than 0.5^o, as this is close to the limit where planes fall over the same voxels for typical MR datasets, as discussed in Section 4.1.

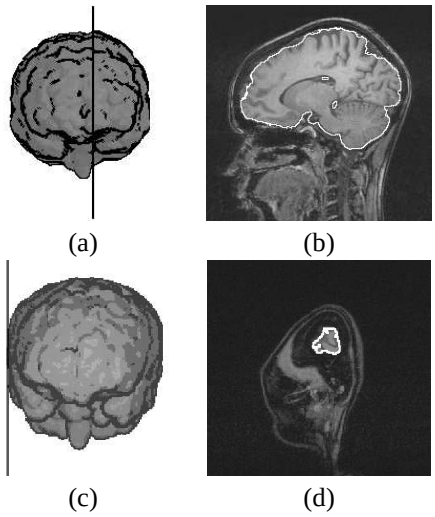


Figure 3: Plane intersection: (a–b) sample plane, brain mask and their intersection (white outline). (c–d) example of a plane tangent to the brain's surface and its small intersection area with the brain mask (delineated in white), overlaid on the original MR image.

4 EVALUATION AND DISCUSSION

4.1 Error Measurement

The discretization of $\mathbb{R}^3$ makes planes that differ by small angles to fall over the same voxels. Consider two planes A and B that differ by an angle Θ (Figure 4). The minimum angle that makes A and B differ by at least 1 voxel at a distance r from the rotation center is given by Equation 2.

$$\Theta = \arctan\left(\frac{1}{r}\right) \quad (2)$$

An MR dataset with 1 mm^3 voxels has a typical maximum dimension of 256 *mm*. For rotations about the center of the volume, the minimum angle that makes planes *A* and *B* differ by at least one voxel within the volume (point p_i in Figure 4) is approximately $\arctan\left(\frac{1}{128}\right) = 0.45^o$. For most MSP applications, we are only concerned about plane differences within the brain. The largest length within the brain is usually longitudinal, reaching up to 200 mm in adult brains. The minimum angle that makes planes *A* and *B* differ by at least one voxel within the brain (point p_b in Figure 4) is approximately $\arctan\left(\frac{1}{100}\right) = 0.57^o$.

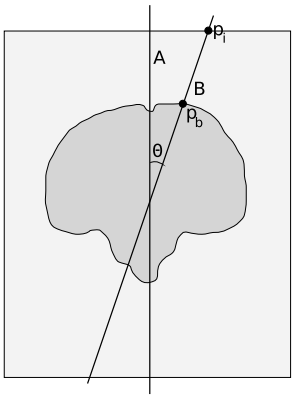


Figure 4: Error measurement in discrete space: points and angles.

Therefore, we can consider errors around 1^o excellent and equivalent results.

4.2 Experiments

We evaluated the method on 64 MR datasets divided in 3 groups: A control group with 20 datasets from subjects with no anomalies, a surgery group with 36 datasets from patients with significant structural variations due to brain surgery, and a phantom group with 8 synthetic datasets with varying levels of noise and inomogeneity, taken from the BrainWeb project (Collins et al., 1998).

All datasets in the control group and most datasets in the surgery group were acquired with a voxel size of $0.98 \times 0.98 \times 1.00$ mm^3. Some images in the surgery group were acquired with a voxel size of $0.98 \times 0.98 \times 1.50\,mm^3$. The images in the phantom group were generated with an isotropic voxel size of 1.00 mm^3. All volumes in the control and surgery groups were interpolated to an isotropic voxel size of 0.98 mm^3 before applying the method.

For each of the 64 datasets, we generated 10 variations (tilted datasets) by applying 10 random transforms composed of translations and rotations of up to 12 *mm* and 12^o in all axes. The method was applied to the 704 datasets (64 untilted, 640 tilted), and visual inspection showed that the method correctly found acceptable approximations of the MSP in all of them. Figure 5 shows sample slices of some datasets and the computed MSPs.

For each tilted dataset, we applied the inverse transform to the computed mid-sagittal plane to project it on its respective untilted dataset space. Thus, for each untilted dataset we obtained 11 planes which should be similar. We measured the angle between all $\binom{11}{2} = 55$ distinct plane pairs. Table 2 shows the mean and standard deviation (σ) of these angles within each group. The low mean angles (column 3) and low standard deviations (column 4) show that the method is robust with regard to linear transformations of the input. The similar values obtained for the 3 groups indicate that the method performs equally well on healthy, pathological and synthetic data. The majority (94.9%) of the angles were less than 3^o, as shown in the histogram of Figure 6. Of $64 \times 55 = 3520$ computed angles, only 5 (0.1%) were above 6^o. The maximum measured angle was 6.9^o. Even in this case (Figure 7), both planes are acceptable in visual inspection, and the large angle between different two computations of the MSP can be related to the non-planarity of the fissure, which allows different planes to match with similar optimal scores. The lower mean angle in the phantom group (column 3, line 3 of Table 2) can be related to the absence of curved fissures in the synthetic datasets. Figure 8 shows some examples of non-planar fissures.

Table 2: Angles between computed MSPs.

Group	**Datasets**	**Angles**	
		Mean	σ
Control	20	1.33^o	0.85^o
Surgery	36	1.32^o	1.03^o
Phantom	8	0.85^o	0.69^o
Overall	64	1.26^o	0.95^o

All experiments were performed on a 2.0 GHz Athlon64 PC running Linux. The method took from 41 to 78 seconds to compute the MSP on each MR dataset (mean: 60.0 seconds). Most of the time was consumed computing the brain mask (stage 1). Stage 1 required from 39 to 69 seconds per dataset (mean: 54.8 seconds), while stage 2 required from 1.4 to 20 seconds (mean: 5.3 seconds). The number of iterations in stage 2 ranged from 0 to 30 (mean: 7.16 iterations).

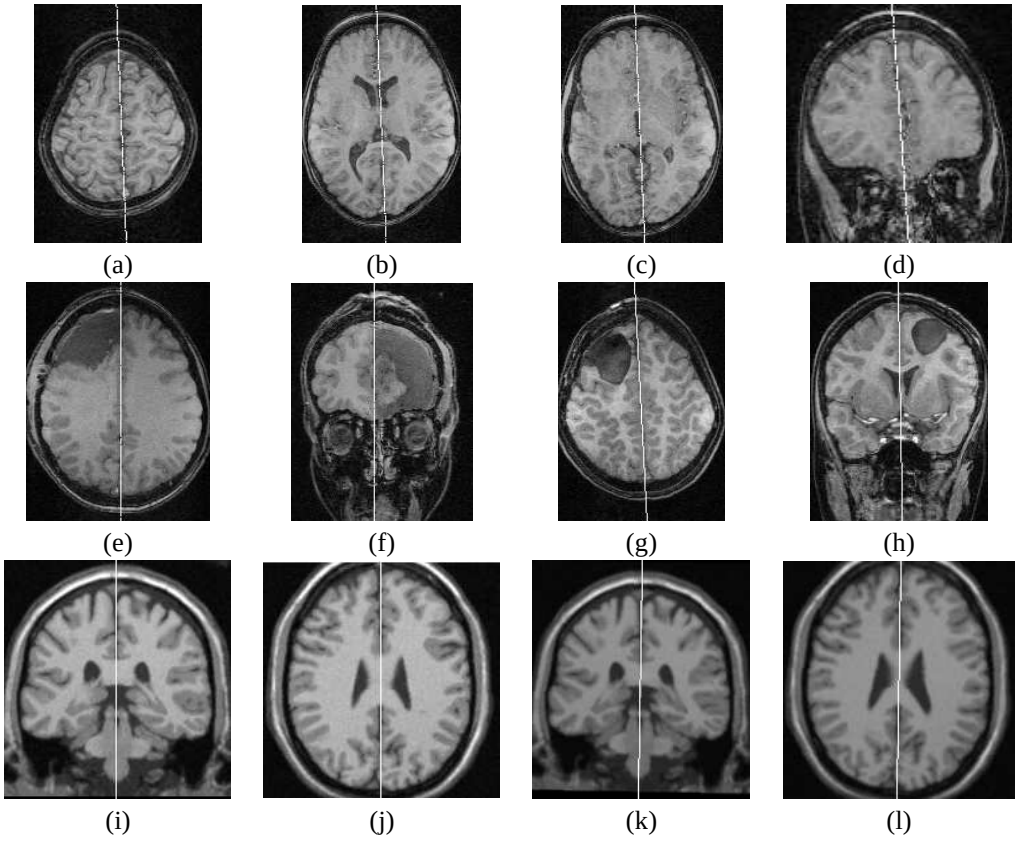


Figure 5: Examples of planes computed by the method: (a–d): sample slices from a control dataset; (e–f) sample slices from a surgery dataset; (g–h) sample slices from another surgery dataset; (i–j): sample slices from a phantom dataset; (k–l): sample slices from a tilted dataset obtained from the one in (i–j).

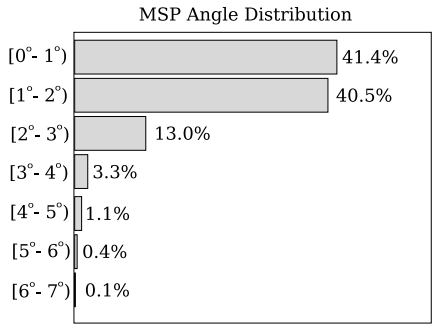


Figure 6: Distribution of the angles between computed mid-sagittal planes.

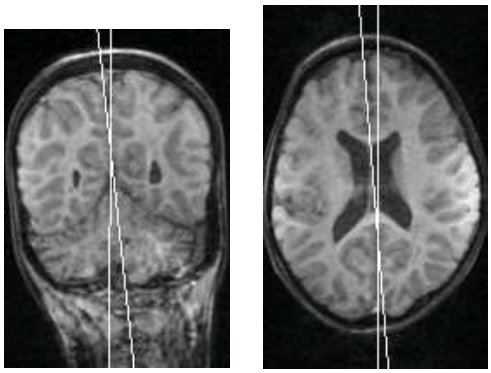


Figure 7: A coronal slice (a) and an axial slice (b) from the case with maximum angular error (6.9^o), with planes in white: The fissure was thick at the top of the head, and curved in the longitudinal direction, allowing the MSP to snap either to the frontal or posterior segments of the fissure, with some degree of freedom.

5 CONCLUSIONS AND FUTURE WORK

We presented a fast and robust method for extraction of the mid-sagittal plane from MR images of the brain. It is based on automatic segmentation of the brain and on a heuristic search based on maximization of CSF within the MSP. We evaluated the method on 64 MR datasets, including images from patients with large surgical cavities (Figure 2a and Figures 5e–h). The method succeeded on all datasets and performed

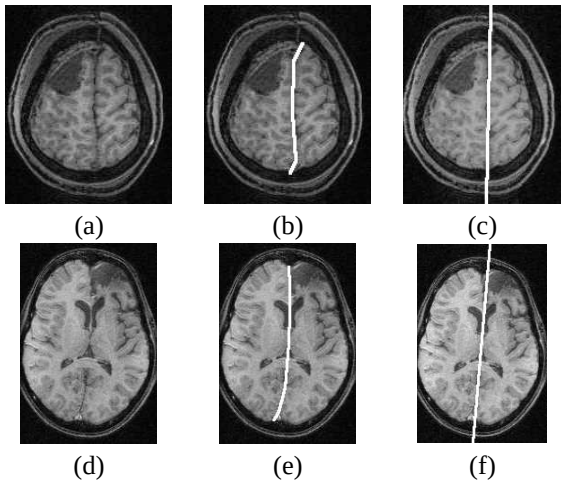


Figure 8: Non-planar fissures: (a) irregular fissure, (b) expert fissure delineation of (a) and (c) MSP computed by our method. (d) Curved fissure, (e) expert fissure delineation of (d) and (f) MSP computed by our method.

equally well on healthy and pathological cases. Rotations and translations of the datasets led to mean MSP variations around 1º, which is not a significant error considering the discrete space of MR datasets. MSP variations over 3º occurred only in cases where the longitudinal fissure was not planar, and multiple planes fitted different segments of the fissure with similar scores. The method required a mean time of 60 seconds to extract the MSP from each MR dataset on a common PC.

Previous fissure-based works were either evaluated on images of healthy patients, on images with small lesions (Volkau et al., 2006), or relied on local symmetry measurements (Hu and Nowinski, 2003). As future work, we intend to implement some of the previous works and compare their accuracy and performance with our method on the same datasets. Brain mask computation is responsible for most of the computing time. We also plan to evaluate how the computation of the brain mask on lower resolutions affect the accuracy and efficiency of the method.

ACKNOWLEDGEMENTS

The authors thank CAPES (Proc. 01P-05866/2007), CNPq (Proc. 302427/04-0), and FAPESP (Proc. 03/13424-1) for the financial support.

REFERENCES

Ardekani, B., Kershaw, J., Braun, M., and Kanno, I. (1997). Automatic detection of the mid-sagittal plane in 3-D brain images. *IEEE Trans. on Medical Imaging*, 16(6):947–952.

Barrick, T. R., Mackay, C. E., Prima, S., Maes, F., Vandermeulen, D., Crow, T. J., and Roberts, N. (2005). Automatic analysis of cerebral asymmetry: na exploratory study of the relationship between brain torque and planum temporale asymmetry. *NeuroImage*, 24(3):678–691.

Bergo, F. P. G., Falcão, A. X., Miranda, P. A. V., and Rocha, L. M. (2007). Automatic image segmentation by tree pruning. Technical Report IC-07-23, Institute of Computing, University of Campinas.

Brummer, M. E. (1991). Hough transform detection of the longitudinal fissure in tomographic head images. *IEEE Trans. on Medical Imaging*, 10(1):66–73.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. (1998). Design and construction of a realistic digital brain phantom. *IEEE Trans. on Medical Imaging*, 17(3):463–468.

Crow, T. J. (1993). Schizophrenia as an anomaly of cerebral asymmetry. In Maurer, K., editor, *Imaging of the Brain in Psychiatry and Related Fields*, pages 3–17. Springer.

Csernansky, J. G., Joshi, S., Wang, L., Haller, J. W., Gado, M., Miller, J. P., Grenander, U., and Miller, M. I. (1998). Hippocampal morphometry in schizophrenia by high dimensional brain mapping. *Proceedings of the National Academy os Sciences of the United States of America*, 95(19):11406–11411.

Csernansky, J. G., Wang, L., Joshi, S., Miller, J. P., Gado, M., Kido, D., McKeel, D., Morris, J. C., and Miller, M. I. (2000). Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus. *Neurology*, 55:1636–1643.

Davidson, R. J. and Hugdahl, K. (1996). *Brain Asymmetry*. MIT Press/Bradford Books.

Falcão, A. X., Bergo, F. P. G., and Miranda, P. A. V. (2004a). Image segmentation by tree pruning. In *Proc. of the XVII Brazillian Symposium on Computer Graphics and Image Processing*, pages 65–71. IEEE.

Falcão, A. X., Stolfi, J., and Lotufo, R. A. (2004b). The image foresting transform: Theory, algorithms and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):19–29.

Geschwind, N. and Levitsky, W. (1968). Human brain: Left-right asymmetries in temporale speech region. *Science*, 161(3837):186–187.

Guillemaud, R., Marais, P., Zisserman, A., McDonald, B., Crow, T. J., and Brady, M. (1996). A three dimensional mid sagittal plane for brain asymmetry measurement. *Schizophrenia Research*, 18(2–3):183–184.

Highley, J. R., DeLisi, L. E., Roberts, N., Webb, J. A., Relja, M., Razi, K., and Crow, T. J. (2003). Sex-dependent

effects of schizophrenia: an MRI study of gyral folding, and cortical and white matter volume. *Psychiatry Research: Neuroimaging*, 124(1):11–23.

Hogan, R. E., Mark, K. E., Choudhuri, I., Wang, L., Joshi, S., Miller, M. I., and Bucholz, R. D. (2000). Magnetic resonance imaging deformation-based segmentation of the hippocampus in patients with mesial temporal sclerosis and temporal lobe epilepsy. *J. Digital Imaging*, 13(2):217–218.

Hu, Q. and Nowinski, W. L. (2003). A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal. *NeuroImage*, 20(4):2153–2165.

Junck, L., Moen, J. G., Hutchins, G. D., Brown, M. B., and Kuhl, D. E. (1990). Correlation methods for the centering, rotation, and alignment of functional brain images. *The Journal of Nuclear Medicine*, 31(7):1220–1226.

Kapouleas, I., Alavi, A., Alves, W. M., Gur, R. E., and Weiss, D. W. (1991). Registration of three dimensional MR and PET images of the human brain without markers. *Radiology*, 181(3):731–739.

Liu, Y., Collins, R. T., and Rothfus, W. E. (2001). Robust midsagittal plane extraction from normal and pathological 3D neuroradiology images. *IEEE Trans. on Medical Imaging*, 20(3):175–192.

Liu, Y., Teverovskiy, L. A., Lopez, O. L., Aizenstein, H., Meltzer, C. C., and Becker, J. T. (2007). Discovery of biomarkers for alzheimer's disease prediction from structural MR images. In *2007 IEEE Intl. Symp. on Biomedical Imaging*, pages 1344–1347. IEEE.

Mackay, C. E., Barrick, T. R., Roberts, N., DeLisi, L. E., Maes, F., Vandermeulen, D., and Crow, T. J. (2003). Application of a new image analysis technique to study brain asymmetry in schizophrenia. *Psychiatry Research*, 124(1):25–35.

Minoshima, S., Berger, K. L., Lee, K. S., and Mintun, M. A. (1992). An automated method for rotational correction and centering of three-dimensional functional brain images. *The Journal of Nuclear Medicine*, 33(8):1579–1585.

Miranda, P. A. V., Bergo, F. P. G., Rocha, L. M., and Falcão, A. X. (2006). Tree-pruning: A new algorithm and its comparative analysis with the watershed transform for automatic image segmentation. In *Proc. of the XIX Brazillian Symposium on Computer Graphics and Image Processing*, pages 37–44. IEEE.

Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:62–66.

Prima, S., Ourselin, S., and Ayache, N. (2002). Computation of the mid-sagittal plane in 3D brain images. *IEEE Trans. on Medical Imaging*, 21(2):122–138.

Smith, S. M. and Jenkinson, M. (1999). Accurate robust symmetry estimation. In *Proc MICCAI '99*, pages 308–317, London, UK. Springer-Verlag.

Styner, M. and Gerig, G. (2000). Hybrid boundary-medial shape description for biologically variable shapes. In *Proc. of IEEE Workshop on Mathematical Methods in Biomedical Imaging Analysis (MMBIA)*, pages 235–242. IEEE.

Sun, C. and Sherrah, J. (1997). 3D symmetry detection using the extended Gaussian image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):164–168.

Talairach, J. and Tournoux, P. (1988). *Co-Planar Steriotaxic Atlas of the Human Brain*. Thieme Medical Publishers.

Teverovskiy, L. and Liu, Y. (2006). Truly 3D midsagittal plane extraction for robust neuroimage registration. In *Proc. of 3rd IEEE Intl. Symp. on Biomedical Imaging*, pages 860–863. IEEE.

Tuzikov, A. V., Colliot, O., and Bloch, I. (2003). Evaluation of the symmetry plane in 3D MR brain images. *Pattern Recognition Letters*, 24(14):2219–2233.

Volkau, I., Prakash, K. N. B., Ananthasubramaniam, A., Aziz, A., and Nowinski, W. L. (2006). Extraction of the midsagittal plane from morphological neuroimages using the Kullback-Leibler's measure. *Medical Image Analysis*, 10(6):863–874.

Wang, L., Joshi, S. C., Miller, M. I., and Csernansky, J. G. (2001). Statistical analysis of hippocampal asymmetry in schizophrenia. *NeuroImage*, 14(3):531–545.

Wu, W.-C., Huang, C.-C., Chung, H.-W., Liou, M., Hsueh, C.-J., Lee, C.-S., Wu, M.-L., , and Chen, C.-Y. (2005). Hippocampal alterations in children with temporal lobe epilepsy with or without a history of febrile convulsions: Evaluations with MR volumetry and proton MR spectroscopy. *AJNR Am J Neuroradiol*, 26(5):1270–1275.

DETERMINE TASK DEMAND FROM BRAIN ACTIVITY

Matthias Honal and Tanja Schultz
Carnegie Mellon University, 407 South Craig Street, Pittsburgh 15213, PA, USA
Karlsruhe University, Am Fasanengarten 5, 78131 Karlsruhe,Germany
honal@ira.uka.de, tanja@{ira.uka.de,cs.cmu.edu}

Keywords: Human-centered systems, Brain Activity, EEG, Task Demand Identification, Meeting and Lecture Scenario.

Abstract: Our society demands ubiquitous mobile devices that offer seamless interaction with everybody, everything, everywhere, at any given time. However, the effectiveness of these devices is limited due to their lack of situational awareness and sense for the users' needs. To overcome this problem we develop intelligent transparent human-centered systems that sense, analyze, and interpret the user's needs. We implemented learning approaches that derive the current task demand from the user's brain activity by measuring the electroencephalogram. Using Support Vector Machines we can discriminate high versus low task demand with an accuracy of 92.2% in session dependent experiments, 87.1% in session independent experiments, and 80.0% in subject independent experiments. To make brain activity measurements less cumbersome, we built a comfortable headband with which we achieve 69% classification accuracy on the same task.

1 INTRODUCTION

Our modern information society increasingly demands ubiquitous mobile computing systems that offer its users seamless interaction with everybody, everything, everywhere, at any time. Although the number and accessibility of mobile devices such as laptop computers, cell phones, and personal digital assistants grows rapidly, the effectiveness in supporting the users to fulfilling their tasks proves to be much smaller than expected. This mainly results from the fact that such devices lack situational awareness and sense for the users' needs. As a consequence users waste their time with manually configuring inflexible devices rather than obtaining relevant information and efficient automatic support to solve their problems and tasks at hand.

It is our believe that the solution lies in intelligent transparent human-centered systems that sense, analyze, and interpret the needs of their users, then adapt themselves accordingly, provide the optimal support to given problems, and finally present the relevant results in an appropriate way. The goal of the work presented here is to solve the analytical part of human-centered systems, i.e. sensing, analyzing, and interpreting the users' needs.

For this purpose we develop learning approaches that derive the users' condition from their brain activity. We are interested in conditions that are important in the context of human-computer interaction and human-human communication. In this particular study we focus on the (mental) task demand as a user condition in the context of lecture presentations and meetings.

The term task demand defines the amount of mental resources required to execute a current activity. Although we are using the general term task demand, we are exclusively concerned about the mental not the physical task demand. Task demand information can be helpful in various situations, e.g. while driving a car, operating machines, or performing other critical tasks. Depending on the level of demand and cognitive load, any distraction arising from electronic devices such as text messages, incoming phone calls, traffic or navigation information, etc. should be suppressed or delayed. Also, the analysis of task demand during computer interaction allows to asses usability. In a lecture scenario, a speaker may use task demand information to tailor the presentation toward the audience.

In this paper we investigate the potential of detecting task demand by measuring the brain activity using scalp electrodes. Although we focus on the system evaluation in the lecture and meeting scenario, the described methods are applicable to any other real-life situation. To make electrode-based recordings acceptable, the following issues must be addressed:

- Robustness: The system needs to be robust against artefacts introduced by speech or body movement
- Usability: EEG sensors and recording device need to be user friendly and comfortable to wear
- Applicability: Measuring brain activity must be feasible in realistic scenarios in real-time.

In this work we are addressing these three goals by relaxing the inconveniences of clinical brain activity recording and make it applicable to real human-computer interaction and human-human communication scenarios.

2 ELECTROENCEPHALOGRAM

The source of the Electroencephalogram (EEG) is neural activity in the cortex, the outmost part of the human brain. This neural activity causes electrical potential differences, which can be measured using scalp electrodes. Information between neurons is transferred via the synapses where chemical reactions take place causing ion movements. These movements result in excitatory or inhibitory electrical potentials in the post-synaptic neurons. The electrical fields emerging from the ion movements are called cortical field potentials and have a dipole structure. If the electrical activity of a huge number of neurons is synchronized, the corresponding dipoles point all in the same direction. Their sum becomes large enough such that potential differences between particular scalp positions and a constant reference point can be measured. EEG characteristics like frequency, amplitude, temporal and topographic relations of certain patterns can then be used to make inferences about underlying neural activities (Zschocke, 1995).

In the EEG which can be measured at the scalp, amplitudes between 1μV and 100μV and frequencies between 0Hz and 80Hz can be observed. These EEG signals show specific characteristics at different scalp positions, depending on the current mental condition. When the human brain is not absorbed by external sensory stimuli or other mental processes, we usually observe the α-activity across the cortex, i.e. rhythmic signals between 8Hz and 13Hz with large amplitudes. When performing higher mental processes the α-activity is attenuated and other activity patterns occur in those cortex regions, where the processes happen. In many cases these patterns are identified by γ-activity, which typically show frequencies around 40Hz and have a lower amplitude than α-activity (Schmidt and Thews, 1997). In this work we assume that the degree of α-activity attenuation and activity at higher frequencies is correlated with task demand. This is justified by the fact that the amplitude of non-α-activity is correlated with the degree of *vigilance*, a physiological continuum between sleepiness and active alertness (Zschocke, 1995). Furthermore, it is known that people are more alert when the task demand is high. The frequency analysis of our recorded data confirms this assumption. During most activity types several cortex regions are involved and task demand is characterized by the *amplitude* of non-α-activity in all regions involved in the current task. This suggests that the activity of the whole cortex must be taken into account to achieve optimal results for task demand estimation.

3 TASK DEMAND & VIGILANCE

A large body of research work concerns the computational analysis of brain activity, applying EEG, functional magnetic resonance imaging, and functional near infrared spectroscopy to areas such as estimation of mental task demand. Several groups reported research on the computational assessment of task demand based on EEG data recorded while varying the task difficulty (Smith, 2001), (Pleydell-Pearce, 2003), (Berka, 2004). These studies focused on the design of intelligent user interfaces that optimize operator performance by adjusting to the predicted task demand level. Regression models were trained to predict task demand from the recorded EEG data. These models used the task difficulty or the rate of errors as references during task execution. The features extracted from the EEG data represented mostly the frequency content of the signals. Positive correlations between predictions and references or predictions and self-estimates of task demand (Smith, 2001) are reported throughout these studies. Pleydell-Pearce (2003) achieved a classification accuracy of 72% for the discrimination of low versus high task demand in subject and session dependent experiments and 71% in subject independent experiments. Task demand assessment has also been done on data from other modalities, including muscular activity (Pleydell-Pearce, 2003), blood hemodynamics (Izzetoglu, 2004), and pupil diameter (Iqbal, 2004). Reasonable results could be achieved with all three modalities. However, correlations between pupil diameter and task demand could only be shown for one interactive task out of a group of various cognitive tasks.

Other work focused on the EEG-based estimation of operator's vigilance during sustained attention tasks (e.g. car driving or operating a power plant). Jung (1997) asked subjects to respond to auditory stimuli which simulate sonar target detection, while EEG was recorded from five

electrodes over the parietal, central and occipital cortex. The error rate in terms of failures to respond to stimuli was then used as reference for a Multi-Layer ANN which was trained with a frequency representation of the EEG signals to predict a vigilance index between 0 and 1. On unknown data a root mean square error (RMS-error) of 0.156 between predictions and references is reported for a subject dependent experimental setup. Duta et al. (Duta, 2004) recorded EEG from the mastoids while subjects had to perform visual attention tasks. Vigilance was labelled by experts who visually inspected the recorded data. Three vigilance categories “alertness”, “intermediate” and “drowsiness” were distinguished. Using the coefficients of an AR model as features for Multi-Layer ANNs 39% to 62% predictions matched the references in subject independent experiments.

4 DATA & METHODS

4.1 Data Capturing

Two different devices were used for data acquisition: an EEG-cap from ElectroCap™ and a self-made EEG-headband (see Figure 1). The majority of data were recorded with the ElectroCap™ using 16 electrodes placed at positions fp1, fp2, f3, f4, f7, f8, fz, t3, t4, t5, t6, p3, p4, pz, o1, and o2 according to the international 10-20 system (Jasper, 1958). Reference electrodes were attached to the ear lobes and linked together before amplification. Although we are aware of the relationship between facial expressions and level of task demand, we decided to exclude the motor cortex from our measurement for two reasons: firstly, the facial muscular activity is partly captured by the frontal EEG electrodes, and secondly we assume that motor activity is of rather minor importance for the assessment of our classification task.

Some data were recorded with a headband, in which we sewed in four electrodes at the forehead positions fp1, fp2, f7, and f8. Reference electrodes were attached to the mastoids and linked together before amplification, the ground electrode was placed at the back of the neck. The headband has three major advantages over the ElectroCap™ which are crucial to real-life applications: the headband is (1) more comfortable to wear, (2) much easier to attach, and (3) better to maintain and clean, also no electrode gel gets in contact with the subject’s hair. The drawback is the limited positioning and number of electrodes compared to the ElectroCap™.

Figure 1: Headband, build-in electrodes at fp1, fp2, f7, f8.

In contrast to recordings for clinical purposes, subjects were allowed to move freely during the recordings to keep the situation as natural as possible, i.e. the subject's head was *not* fixated. Consequently we had to deal with data artefacts introduced by muscular activity (some recordings required speaking as well). Strategies to remove those artefacts will be described in section 4.2.

Amplification and A/D-conversion was done with a 16 channel VarioPort™ physiological data recorder (Becker, 2005). Each channel had an amplification factor of 2775 and a frequency range from 0.9Hz to 60Hz. After amplification, A/D conversion was performed using 4096 A/D-steps and a sampling rate of 256 Hz. The data was transferred instantaneously from the amplifier to a computer via an RS232 port for online processing. The port capacity is limited to 115200 Bits per second which corresponds to 28 electrode channels at a sampling of 256Hz. Although sampling with a lower frequency should be sufficient to avoid aliasing when considering the amplifier’s upper cutoff-frequency of 60Hz, we decided to go with 256Hz since for technical reasons the slope of the band pass filter is very small.

4.2 Data Preprocessing

Figure 2 summarizes the signal processing steps of our task demand estimation system. After EEG recording, artefacts are removed applying independent component analysis (ICA). A short time Fourier transform (STFT) is used for feature extraction. After feature normalization and averaging over temporally adjacent features, different methods for reducing the dimensionality are used. Finally, Support Vector Machines (SVMs) or Artificial Neural Networks (ANNs) for classification or regression are applied to obtain task demand predictions. We also applied Self-Organizing-Maps (SOMs) to determine which levels of task demand can be reliably discriminated.

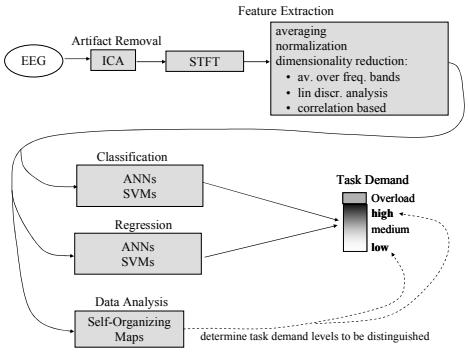


Figure 2: Task Demand Estimation System.

4.2.1 Artefact Removal

Artefacts such as muscular activity and especially eye movements contaminate the EEG signal, since the corresponding electrical potentials are an order of magnitude larger than the EEG sources. This causes in particular problems in the EEG that is measured over the frontal cortex. ICA has shown to be very efficient for artefact removal in EEG data (Jung et al., 2000).

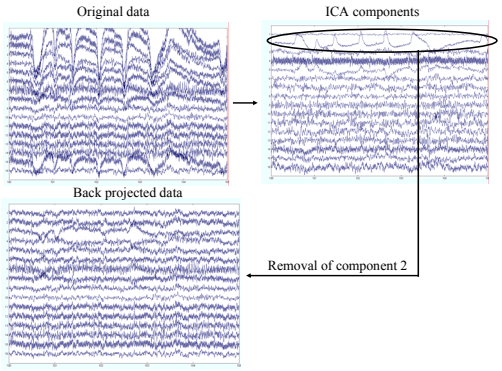


Figure 3: Artefact removal applying ICA: (1) independent components are computed from the original data (top left), (2) the second component (eye blinking artefact) is identified and rejected (top right), and (3) the data is projected back to the original space (bottom left).

To apply ICA to EEG data it is assumed that the signal measured at one electrode can be described by a linear combination of signals emerging from independent processes (i.e. cortical field potentials, muscular artefacts, 60Hz AC noise): Let $\mathbf{x}(t)$ be the vector of signals measured at all electrodes at time t and $\mathbf{s}(t)$ be the independent components. Then $\mathbf{x}(t)$ can be expressed by $\mathbf{x}(t) = A \cdot \mathbf{s}(t)$, where A is called mixing matrix. ICA computes the matrix A, or its inverse the de-mixing matrix W, such that independent components can be estimated from the measured signals (Hyväarinen et al., 2000). Artefact components can then be identified either by visual inspection of the training data or by using cross-validation and be rejected from the data. The remaining components are projected back into the original coordinate system (see Figure 3). For ICA computation we used the open source Matlab toolbox EEGLAB (Delorme et al., 2004), which applies the Informax algorithm to the matrix estimation.

4.2.2 Feature Extraction, Averaging and Normalization

After artefact removal we computed the power spectrum of the time signal applying STFT. For two-second long segments overlapping by one second, features were computed representing the content of frequency bands with 0.5Hz width. This results in one feature vector per second. The dimensionality of one feature vector for 16 electrode channels and frequencies ranging from 0 to 45Hz is 16·90=1440. To reduce the influence of outliers final feature vectors for each time point were obtained by averaging over *k* previous features. To compensate for different ranges in the frequency bands, we normalized each feature using the following two normalization approaches:

- **GlobalNorm:** Feature means and variances are calculated based on the complete training set. Calculated values are used globally for mean subtraction and variance normalization on all data (training, validation, and test data).
- **UserNorm:** Feature means and variances are calculated on training, validation, and test data separately for each user. Then, user-specific mean subtraction and variance normalization is applied.

4.2.3 Feature Reduction

Since the dimensionality of the feature vector may be large compared to the amount of training data, we investigated various feature reduction methods. A straightforward approach is to average over adjacent frequency bands, another approach is the Linear Discriminant Analysis (LDA), which selects features according to their discriminative power (Fukunaga, 1972). For sparse data and large dimensionalities, LDA estimation may become ill-conditioned. Therefore, we also applied a correlation-based feature reduction method, which selects those features that correlate best with the variable to be predicted. This method proved to be particular useful for the assessment of task demand, since – in contrast to LDA – it takes into account the continuous nature of the predicted variable.

4.3 Data Analysis

To learn more about the data structure and to gain insights into the granularity and distinctness of task demand levels, we generated self-organizing maps (SOMs) (Kohonen, 1995) for the training data. After obtaining the Best Matching Unit (BMU) for each training example, a map was calculated which visualizes colour-coded clusters corresponding to different task demand levels. Thus the spatial relation between the feature vectors belonging to the different task demand levels can be visualized concisely on a two dimensional grid. Although the SOM-based analysis may indicate which task demand levels are easy to discriminate, the hypotheses have to be verified experimentally on test data. SOM training and visualization were performed with the MATLAB™ based SOM-Toolbox (Vesanto et al., 2000).

4.4 Learning Methods

We investigated two types of classifiers: Multilayer ANNs and SVMs. ANN classifiers were trained with standard back-propagation, based on feed-forward networks with a `tanh` activation function and one hidden layer. For all ANNs early stopping regularization was performed and the number of neurons in the hidden layer was determined on the validation data. For SVM-based classification we used an implementation of SVM^{light} (Joachims, 1999), which directly addresses the multi-class problem (Tsochantaridis, 2004). SVMs were restricted to linear kernels to limit computational costs and avoid extensive parameter tuning. By treating the task demand levels as class labels (e.g. “low”, “medium”, “high”), both classification methods can be applied to the problem of task demand estimation. To exploit the information contained in the ordinal scaling of the different class labels, we investigated the regression versions of ANNs and SVMs as well.

Since ANN predictions fluctuate due to random weight initializations, predictions from five networks trained on the same data were combined using majority decisions (in case of classification) or averaging (in case of regression).

4.5 Evaluation Methods

The system performance for task demand assessment is evaluated in terms of classification accuracy. When regression methods are used, class labels are assigned numeric values and each prediction is assigned to the label with the closest value. Although confusion matrices could lead to a deeper understanding of pros and cons of the prediction methods, we decided to use the more concise classification accuracies. Results presented here are averages over all test sets and all class accuracies. The latter gives more reliable results in the presence of unbalanced test sets.

We use the normalized expected loss to compare accuracies that were calculated based on different numbers of classes. Comparing accuracies directly would not be appropriate since the chance accuracy $A^{(c)}$ varies with the number of classes. The normalized expected loss relates the observed error to the chance error and thus makes it independent from the number of classes. The value of the normalized expected loss is bound by $1/ A^{(c)}$ and ranges between 0 and 1.

5 EXPERIMENTS

We conducted various experiments to evaluate task demand assessment and collected EEG data for this purpose, using both the headband and the ElectroCap™. In offline experiments we analyzed and optimized the processing steps of the system.

5.1 Data Collection

Task demand data was collected from subjects perceiving an audio-visual slide presentation. The presentations were tailored to the subjects’ educational background and designed to provoke each task demand level with equal amount of time. The presentations were video-taped so that each of the subjects could evaluate their task demand afterwards by watching the tape. We defined the following task demand levels:

- Low: All details of the presentation are well understood with low mental effort.
- Medium: Some mental effort is required to follow the presentation, not all details may be understood.
- High: All available mental resources are required to understand at least the essence of the topic. Most of the details are not understood.
- Overload: The presentation topic is not understood. The subject is overwhelmed, disengaged and makes no more effort to understand the presentation.

In total 7690 seconds of data were recorded with the ElectroCap™ from six students (three male, three female) between 23 and 26 years old. One subject was recorded twice. 1918 seconds of data were recorded with the headband from two students (one male, one female) between 21 and 28 years old.

5.2 Experimental Setup

One major goal of our experiments was to investigate the impact of user and session dependencies on the system performance. The other goal was to examine the efficiency and performance of the headband compared to the ElectroCap™. We therefore conducted user/session dependent and independent experiments on ElectroCap™ and headband recordings using the following setup:

UD: **U**ser and session **d**ependent setup: Different subsets of the same session were used for training (80%), validation (10%), and testing (10%). Four sessions were recorded with the ElectroCap™ and two with the headband.

UI: **U**ser and session **i**ndependent setup: The system was trained on three of the four ElectroCap™ recording sessions and tested on the fourth session in a round-robin fashion. For better comparability the same test sets as for setup UD were used. Validation was performed on two held-out ElectroCap™ recording sessions.

SI: **S**ession **i**ndependent but user dependent setup: One subject was recorded twice in two separate sessions using the ElectroCap™. The system was trained on one session and tested on the other, without validation set.

5.3 Results – Data Analysis

Figure 4 compares for one subject the SOM trained on all task demand levels (left-hand side) to the SOM trained on high and low task demand level (right-hand side). The grey-scaled dots represent the best matching units (BMUs) on the grid belonging to the feature vectors of different task demand levels. The size of the dots is proportional to the amount of feature vectors that share the same BMU. Obviously we see a large overlap between the BMUs when all four task demand levels are considered, while the BMUs for low and high task demand seem to be well separable. Same observations were made for the SOMs trained on other subjects.

Baseline results on the UD setup (no averaging, **GlobalNorm** normalization, no feature reduction, linear classification SVMs) confirmed our expectation that the four task demand levels are difficult to discriminate (classification accuracy 40%, normalized expected loss 0.81). When distinguishing low versus high task demand we achieved a classification accuracy of 78% and a normalized expected loss of 0.43. The major reason for the poor results on discriminating all four levels is that subjects had difficulties to identify the boundaries between adjacent demand levels. To investigate this we asked some subjects to re-evaluate their task demand at a later time. We found a low intracoder agreement among adjacent task demand levels, while high versus low task demands were rarely confused. In the remainder of this section we will therefore focus on the discrimination between low and high task demand.

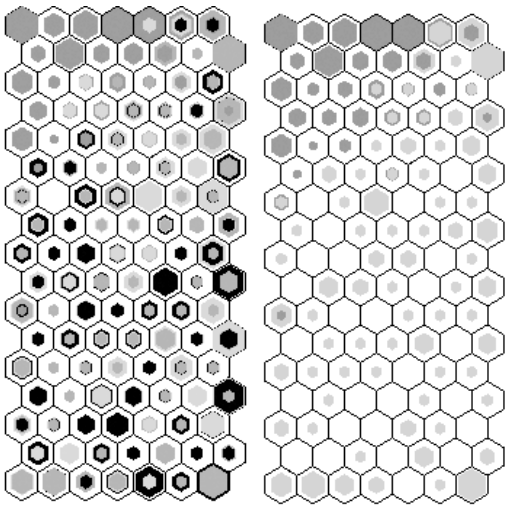


Figure 4: SOM trained on all four task demand levels (left-hand side) and on low vs. high task demand (right-hand side). Grey scale intensity indicates task demand level, ranging from low (light) to overload (dark).

Table 1 shows the average amount of data per subject after removing the medium and overload task demand recordings.

Table 1: Data per subject (in seconds) for all setups.

Setup	Training	Validation	Test
UD	247	31	64
UI	740	229	64
SI	257	-	48

5.4 Results – Learning Method

Table 2 compares the regression and classification versions of ANNs and SVMs for the baseline system (no averaging, **GlobalNorm** normalization, no feature reduction). For all three experimental setups SVMs perform better than ANNs. For setup SI the regression SVMs significantly outperform the classification SVMs. For the other setups the differences between the two SVM variants are rather small. Since at least theoretically the regression SVMs should be able to better exploit the ordinal scaled information given in the task demand levels, we decided to use these in the remainder of our experiments.

Table 2: Baseline system performance for all setups; classification ([c]) and regression methods ([r]); In parentheses: standard deviation for five ANN experiments.

Setup	UD	UI	SI
SVM c	**81%**	72%	66%
SVM r	79%	74%	**73%**
ANN c	78% (7%)	70% (3%)	53% (5%)
ANN r	71% (3%)	69% (3%)	66% (5%)

5.5 Results – Normalization and Feature Reduction

In the following experiments we optimized the processing steps of our system in a greedy fashion on the validation set. Table 3 shows the classification accuracies for all experimental setups with the optimal parameters (given in parentheses).

Averaging over k=2 feature vectors improved the results for the UD and UI setup. The use of normalization method **UserNorm** instead of the baseline method **GlobalNorm** improved results for setups UI and SI. This matches our expectation, since this method reduces the variability across sessions (UI and SI) as well as across subjects (UI). Normalization is not relevant for the user dependent setup (UD) since it only applies when data of different subjects are used for training and test.

Table 3: Results for the optimized task demand system.

Setup	UD	UI	SI
Baseline	78%	74%	73%
Averaging (k=2)	82%	79%	73%
Normalizing (UserNorm)	N/A	**80%**	**87%**
Feature Reduction (Corr-based)	**92%**	77%	66%

Feature reduction was only successful for UD, where a correlation based reduction from 1440 to 80 features yielded considerable improvements. For the other setups feature reduction did not help, probably since despite normalization the data variability was too large. Consequently, features which were well correlated with task demand on the training data exhibited poor correlations with task demand on the test data. Comparing the results of feature reduction among the different setups is difficult since the optimal number of 80 features for the UD setup was determined on the validation set, while we set this number manually to 240 for the SI and UI setup as the validation method did not give any reasonable optimum.

Averaging over adjacent frequency bands for feature reduction corresponds to putting features into bins of size b. We observed that even for large numbers of b the results did not drop much for any of the setups. For b=45 (two features per electrode, i.e. lower and the upper frequencies) results are in the same range as without feature reduction. For b=90 (one feature per electrode, 8 features in total) results dropped significantly. This suggests that it is sufficient to consider for task demand estimation the content of two broad frequency bands: the lower frequencies (around the α-band) and the higher frequencies (around the γ-band). Experiments to investigate this hypothesis are planned. The feature reduction would benefit from more reliable model estimation and reduced computational costs.

5.6 ElectroCap™ versus Headband

After optimizing the system parameters, experiments using the UD setup were conducted on the headband data. A classification accuracy of 69% could be achieved. This compares to 69% using the four ElectroCap™ recordings with 4 electrodes and 82% with 16 electrodes. These results were achieved without correlation based feature reduction. For the reduced number of electrodes, the classification accuracies for half of the subjects are at least 86% or better, while for the other half they are around chance. This implies that the feasibility of task demand estimation based on four electrodes might depend on the subject or even on the presentation itself. As described above the presentations and topics were tailored towards the educational background of the subjects.

6 CONCLUSIONS

In this paper we described our efforts in building human-centered systems that sense, analyze, and interpret the users' needs. We implemented several learning approaches that derive the task demand from the user's brain activity. Our system was built and evaluated in the domain of meeting and lecture scenarios. For the prediction of low versus high task demand during a presentation we obtained accuracies of 92% in session dependent experiments, 87% in subject dependent but session independent experiments, and 80% in subject independent experiments. To make brain activity measurements less cumbersome, we built a comfortable headband with which we achieved 69% classification accuracy for low versus high task demand discrimination. Based on our findings we developed an online system that derives user states from brain activity using the headband (Honal et al., 2005). A screenshot of our prototype is shown in Figure 5.

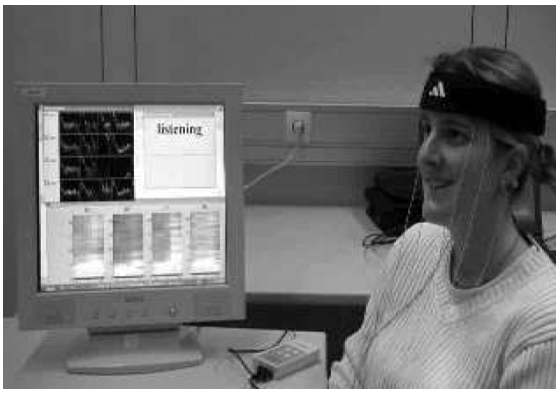


Figure 5: Screenshot of our prototype online brain activity system. The upper left monitor area displays the EEG signal; the hypothesized current user state is shown in the upper right corner. Spectrograms for the headband electrodes fp1, fp2 f7 and f8 are shown at the bottom.

ACKNOWLEDGEMENTS

This material is in part based upon work supported by the European Union (EU) under the integrated project CHIL (Grant number IST-506909). Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the EU. The authors would like to thank Laura Honal and Dana Burlan for their cooperation and patience during the data collection, Christian Mayer and Markus Warga for implementing the recording software and tools, as well as Klaus Becker and Gerhard Mutz for their support with respect to the recording devices. Without the contributions of all these people, this work would not have been possible.

REFERENCES

Becker, K. VarioPort™. http://www.becker-meditec.com/.

Berka, C., Levendowski, D., Cvetinovic, M., et al., 2004. Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. Int. Journal of HCI, 17(2):151–170.

Delorme A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. Journal of Neuroscience Methods, 134:9-21.

Duta, M., Alford, C., Wilon, S., and Tarassenko, L., 2004. Neural Network Analysis of the Mastoid EEG for the Assessment of Vigilance. Int. Journal of HCI, 17(2):171–195.

Electro-Cap™, Electro-Cap International, Inc.: http://www.electro-cap.com/.

Fukunaga, K., 1972. "Introduction to Statistical Pattern Recognition", Academic Press, New York, London.

Honal, M. and Schultz, T., 2005. Identifying User State using Electroencephalographic Data, Proceedings of the International Conference on Multimodal Input (ICMI), Trento, Italy, October 2005.

Hyväarinen, A. and Oja, E., 2000. Independent Component Analysis: Algorithms and Applications. Neural Networks, 13(4-5):411-430.

Iqbal, S., Zheng, X., and Bailey, B., 2004. Task evoked pupillary response to mental workload in human computer interaction. In Proceedings of Conference of Human Factors in Computer Systems (CHI).

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., and Chance, B., 2004. Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of HCI, 17(2):211–227.

Jasper, H. H.. 1958. The ten-twenty electrode system of the International Federation. Electroencephalographic Clinical Neurophysiolgy, 10:371–375.

Joachims, T. (1999). Making Large-Scale SVM Learning Practical, chapter 11.MIT-Press.

Jung, T., Makeig, S., Humphries, C., Lee, T., Mckeown, M., Iragui, V., Sejnowski, T. (2000) Removing Electroencephalographic Artifacts by Blind Source Separation. Psychophysiology, 37(2):163-17

Jung, T.P., Makeig, S., Stensmo, M., and Sejnowski, T.J., 1997. Estimating Alertness from the EEG Power Spectrum. IEEE Transactions on Biomedical Engineering, 4(1):60–69, January

Pleydell-Pearce, C.W., Whitecross, S.E., and Dickson, B.T.. 2003. Multivariate Analysis of EEG: Predicting Cognition on the basis of Frequency Decomposition, Inter-electrode Correlation, Coherence, Cross Phase and Cross Power. In Proceedings of 38th HICCS

Schmidt, F. and Thews, G. (editors) (1997). Physiologie des Menschen. Springer

Smith, M., Gevins, A., Brown, H., Karnik, A., and Du, R., 2001. Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction. Human Factors, 43(3):366–380.

Tsochantaridis, I., Hofmann, T., Joachims, T., and Altun, Y. (2004). Support Vector Machine Learning for Interdependent and Structured Output Spaces. In Proceedings of the ICML.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). SOM Toolbox for Matlab 5. Technical report, Helsinki University of Technology.

Zschocke, S. (1995). Klinische Elektroenzephalographie. Springer

A PROBABILISTIC TRACKING APPROACH TO ROOT MEASUREMENT IN IMAGES

Particle Filter Tracking is used to Measure Roots, via a Probabilistic Graph

Andrew French [1], Malcolm Bennett [1], Caroline Howells [1], Dhaval Patel [2] and Tony Pridmore [1]
[1] *Centre for Plant Integrative Biology, University of Nottingham*
Sutton Bonington Campus, Loughborough,UK. LE12 5RD
andrew.french@cpib.ac.uk
[2] *Plant Sciences, School of Biosciences, University of Nottingham*
Sutton Bonington Campus, Loughborough,UK. LE12 5RD

Keywords: Particle filter, image analysis, quantification, bioinformatics.

Abstract: This paper introduces a new methodology to aid the tracing and measurement of lines in digital images. The techniques in this paper have specifically been applied to the labour intensive process of measuring roots in digital images. Current manual methods can be slow and error prone, and so we propose a semi-automatic way to trace the root image and measure the corresponding length in the image plane. This is achieved using a particle filter tracker, normally applied to object tracking though time, to trace along a root in an image. The samples the particle filter generates are used to build a probabilistic graph across the root location in the image, and this is traversed to produce a final estimate of length. The software is compared to real-world and artificial length data. Extensions of the algorithm are noted, including the automatic detection of the end of the root, and the detection of multiple growth modes using a mixed state particle filter.

1 INTRODUCTION

Within biological science experiments it is common for measurements of samples of interest to be made from digital images. This paper is concerned in particular with the length measurement of roots of *Arabidopsis thaliana* from images of plates of roots taken with a digital camera. This process is largely carried out manually, by measuring the roots by hand in an image processing package such as the public domain ImageJ (Abramoff et al., 2004). For each root, the user must manually mark a line along its length, and the software then calculates the length. Other methods measure mouse travel distance as the user traces an image of a root (Pateña & Ingram, 2000). Clearly, it would be useful to automate as much of this process as possible, particularly the laborious and error-prone manual tracing step.

Some tools already exist to aid with root measurement, but each has its drawbacks or specific mode of operation. RootLM (Qi et al., 2007), for example, is capable of measuring growth rates over daily intervals, but requires root growth to be marked up on the petri dish in marker pen, and the removal of the actual roots, prior to scanning. MR-RIPL 2.0 (Smucker, 2007) estimates the lengths and widths of roots by applying global thresholding and thinning processes to identify roots on an opposing intenisty background, an approach which can be hampered by clutter on the image plane. Other tools similarly use thresholding and thinning to isolate the roots (Bauhus & Messier, 1999), and can also be sensitive to noise and clutter.

In this paper, a robust probabilistic method of root length measurement is presented. This approach uses a particle filter to track along the root image, building a probabilistic graph using the sample locations and observed likelihoods at those locations. The graph is then pruned, removing low probabilty vertices, and a shortest path algorithm is applied to describe the line down the centre of the root. This line can then be used to provide a measurement of root length. The approach is found to work well, handling images with clutter and lighting variations obscuring parts of the root.

Section 2, describes how the shape of the root is traced and how from this a measurement of length is

calculated. Results are presented in Section 3 which compare this new algorithm to manual methods on real-life and synthetic images. The discussion in Section 4 then examines the results, and an appraisal of the algorithm is presented, including the possiblity of wider applications.

2 METHOD

2.1 Root Tracing

Before a quantification of root length can take place, an accurate tracing of the root image is required. The approach adopted here is based on a particle filter tracking technique. Particle filtering, first developed as a method of tracking moving objects through an image sequence, is a way of representing system states that might not be definable with closed-form functions. States are represented using probability density functions (PDFs), or rather discrete estimates of them modelled by particle sets. A particle set can represent a function by sampling the distribution and weighting particles corresponding to these samples. Contained within a particle is all the information about the state of the system at that time, for example, for target tracking across the image plane a particle might contain (x, y) coordinates and velocity information.

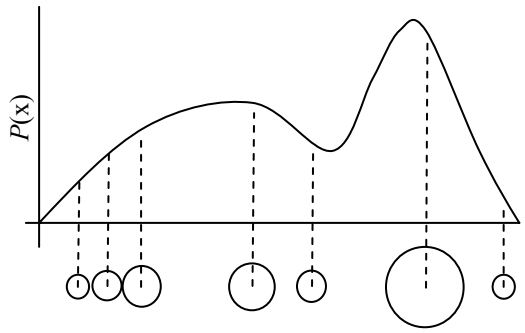


Figure 1: Representing a continuous PDF using a particle set of 7 particles. The particles are randomly distributed, and the weight of a particle (represented by the size of the circle) corresponds to the value of the function at that point.

As shown in Figure 1, a continuous function can be approximated by a finite number of particles and their weights. The more particles that are used, the more accurate the representation. Normally, when tracking a moving object, the PDF is a measure of the probability of the target actually being at a position, and is measured using an observation model which reports high probabilities when it is over an area of image that matches the target appearance model. So, in Figure 1 above, the first, lower peak might represent the location of some background clutter, and the second, higher peak the actual target.

When tracking a target over time, the predicted position of the object depends on both where the object was at the last timestep, and on a motion element determined by a dynamic model of the target. Propagating the continuous PDF estimate of position forward in time with this motion model tends to shift the curve in the direction of the prediction. Adding an additional random diffusion term, simulating noise in the tracking, has the effect of smoothing the PDF, and after the motion phase, the PDF is reinforced with measurements using the observation model.

In the discrete case, where a finite set of particles represents the distribution, a set of particles are selected and have the motion model applied to their state. These particles are selected with a probability in proportion to their weight, and are replaced after selection, ready for re-selection. This has the effect of generating a new particle set in which the particles tend to cluster mainly around the higher probability peaks, with fewer particles representing the lower probability valleys. As the peaks are what we are interested in (they suggest where our target actually is), this importance sampling improves tracking performance.

This process is known as factored sampling. Every time we select a particle we process its state parameters forward in time using the motion model, and then weight this particle based on the observation model at this new position. This gives us our new set of weighted particles, ready for another iteration of the algorithm. One of the attractions of particle filtering methods is that the sample set size remains constant, so the algorithm runs in a predictable time, and the quality of the representation of the PDF can be increased by increasing *N*, the number of particles in the set. A classic example of a computer vision tracking algorithm which uses an algorithm like this is the Condensation algorithm (Blake & Isard, 1998).

We have adapted this tracking model so that instead of being used over *time*, it is used over *space*, to trace along a root in a digital image. It is assumed that the root lies approximately parallel to one of the major image axes, so we know approximately which way to trace the image. We shall assume here that the root lies approximately parallel to the y-axis.

The algorithm proceeds as follows:

1. The user selects, in the image plane, the starting point of the root to be traced. Around this an initial distribution of *N* particles is built. This distribution is normally a Gaussian distribution along the x-axis, centred on the user's click point. The y-locations are fixed to the user's set y-coordinate for reasons which will become clear. Initially all these particles are given equal probability weights.
2. Particles are selected with replacement in proportion to their probability weighting. As each particle is selected, its y-coordinate is incremented by exactly 1 pixel, and the x-coordinate is processed forward using its predictive 'motion' model plus a small level of random Gaussian noise.
3. The new particles are weighted by comparing the image at their current location with the observation model of a root cross section.
4. The probabilities associated with each particle, and the locations of the particles are stored as nodes in a graph – this will be used later on. All the nodes at time *t* are connected to all the nodes at time *t-1*, therefore each iteration *N* new nodes and *N*N* new edges are added to the graph.
5. The algorithm repeats to step 2, until the root is fully traced and the user stops the process at iteration *I*.

Fixing the y-coordinate to proceed at an increment of 1 pixel per iteration provides an external force to the tracing algorithm to move the trace down the root by exactly one pixel at a time. This is analogous to tracing a line by hand using a pencil, starting at one end and moving smoothly to the other. This external force along the *y* axis, combined with the motion model to cope with curvature along the root in the *x* axis, replaces the motion model used when tracking moving objects, and allows an uninterrupted and unrepeated line to trace along the root.

At the completion of the algorithm, there exists a graph *G* with *N*I* nodes and *N*N*I* edges. Each node represents a weighted sample from the particle filter, and has a corresponding weight (probability), and coordinate within the image plane. An example visualization of how the graph relates to the particles and image is presented in Figure 2.

It should be noted that currently the tracing is ended manually by the user when the trace is seen to reach the end of the root. Detecting when tracking should cease is a hard problem as tracking algorithms assume the target to exist at the next timestep. The authors are working on a robust method to detect the end of the root automatically, which is mentioned further in Section 4.1.

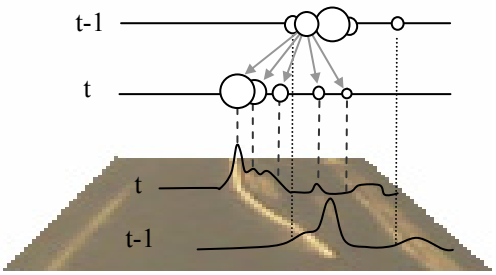


Figure 2: Illustration of the relationship between the images, observation model output (curves), particle weights (circles), and graph connections between two steps in the algorithm, *t* and *t-1*. Note some of the lines connecting the curves to particle weights at *t-1* have been omitted for clarity. Grey arrows indicate the edges from *one* particle when it is mapped to a node in the graph– in fact *every* node at each layer has edges connecting to all nodes generated at the next step of the algorithm.

2.2 Probabilistic Graph

Graph *G* can be thought of as a 3D surface map which represents probabilities associated with each possible root location. Using this we aim to produce an accurate measure of root length. This is done by removing low probability nodes from the graph and then finding the minimum distance route through the remaining graph, from the start position to an end node.

Low probability nodes are removed as they represent areas of the image space explored by the particles which are not centred over the root. During the root tracking procedure, at each iteration particles are spread around the root width, to increase the chances of finding the root at each step. The aim of our graph pruning procedure is to remove those nodes from the graph that represent locations in the image which are so far from the ideal observation model result that they cannot represent a root location. This information is useful during the online tracking, but is not needed for the offline graph traversal.

To do this pruning, we simply remove probabilities whose measurements fall below a certain number of standard deviations from the mean measurement across the set, although any heuristic-based method could be employed here to remove low quality nodes from the graph. This has the effect of producing a leaner graph which only covers the space occupied by the root.

To actually find the shortest path through the graph, Dijkstra's method for determining shortest

paths was implemented. This method involves a greedy algorithm which determines the shortest distance to each node as it traverses the graph, in our case along the length of the root, therefore giving the shortest path along the length of the root, through the remaining high-quality nodes.

3 RESULTS

The proposed method was tested by comparing measurements of roots obtained using standard manual techniques and using the new software. The particle filter used 25 particles in all the tests.

3.1 Software versus Non-expert User

The software was tested with an image of plated roots (Figure 3). The aim was to measure the length of the roots from the black line to the root tip. The image had been taken with an off the shelf digital camera, and was stored in a compressed JPEG format, at a resolution of 783x576 for the close-up in Figure 3. The roots were measured manually, by an inexperienced user, using the measure tool in ImageJ (Abramoff et al., 2004). This measurement was repeated 5 times. The particle-filter software was also run five times. An example output is presented in Figure 3, while numerical results are given in Table 1.

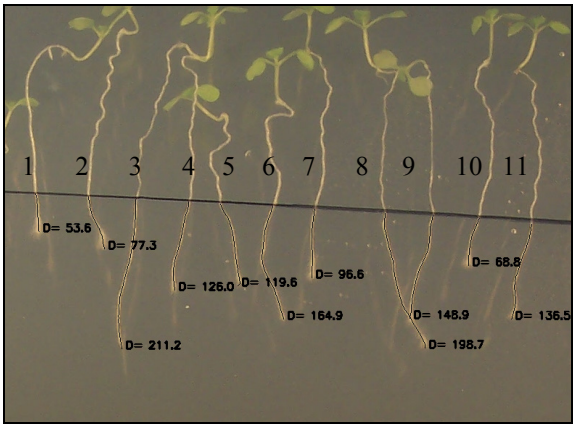


Figure 3: Image of growing roots with the software output overlaid. The root numbers refer to the results in Table 1.

Table 1: Results of a comparison between the new software and manual measurements made by a non-expert.

Root	Error between means, pixels	Relative error (Mean-mean error as % of manual measure)
1	-0.1	0.18
2	1.26	1.65
3	3.16	1.52
4	1.3	1.02
5	2.14	1.84
6	2.12	1.29
7	3.86	4.1
8	4.36	2.23
9	1.58	1.05
10	1.98	2.96
11	1.84	1.35

The mean length for these roots is 126.4 pixels, from the ground truth. The average standard deviation for the manual measures was 1.97 pixels, and 1.71 pixels for the proposed software.

The average time taken to measure manually the roots on the plate in Figure 3 once each was 112 seconds. The new software, including the time for user interactions clicking on the image and stopping the tracing, took 70 seconds.

The average relative error from Table 1 is 1.7%. Root 7 produces the most ambiguous measures from the new software, but on inspection its root tip is blurry and ambiguous in the image itself, which may explain the error. This situation might produce measurements with high variability when different subjects are asked to perform the measurement manually.

3.2 Software versus Expert User

The software was also tested against manual measurements made by an expert user. The root images used in this section are more complex, with the roots showing many lateral roots. There are also significant reflections from the rear of the plate, and the images are of low resolution (640x480), all of which makes this scenario a challenge for the software.

For this experiment, five roots in Figure 4 were manually measured in ImageJ by a trained biologist familiar with making such measurements. This measurement was compared with the average results of five runs of the new software approach. The data is presented in millimetres; using the ruler in Figure 3 a conversion was calculated between pixels and millimetres. The results are presented in Table 2. The test image is presented in Figure 4.

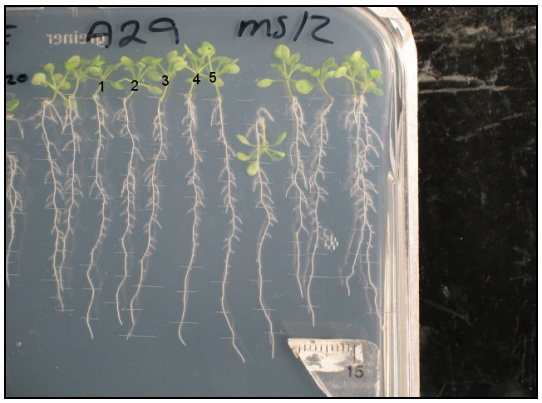


Figure 4: Image of growing roots. Note the large numbers of lateral roots and reflections which clutter the image. The root numbers refer to the results in Table 2.

Table 2: Results of a comparison between the new software and manual measurements made by an expert user.

Root	Error between means, mm	Relative error (Mean-mean error as % of manual measure)
1	0.78	1.67
2	-1.95	4.54
3	1.32	2.89
4	0.86	1.81
5	0.02	0.05

The average root length from the manual measures was 46.8mm. The proposed software measures had an average standard deviation of 0.23mm. The average relative error from Table 2 is 2.2%.

The automatic tracing of roots 1 and 2 suffered the most due to interference from the lateral roots. Examples of such error cases are presented in Figure 5.

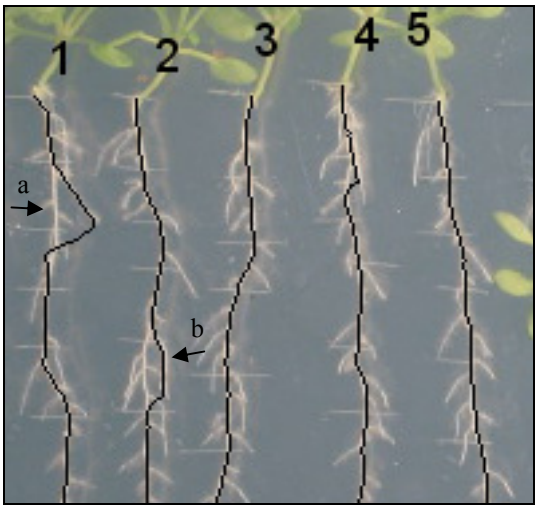


Figure 5: Example output, with error cases marked.

Figure 5 illustrates two of the most common error cases. For case (a), the lateral root is followed rather than the main root about 50% of the time. This is because when tracing the line, the tracking algorithm reaches a junction, and as the motion model predicts the line to continue roughly half way between the two actual lines, and both lines produce very similar measurement models, half the time the algorithm will take one route, and the other half of the time the other route will be followed.

The particle filtering trackers can cope with this kind of ambiguity over short distances, but over longer distances the samples tend to all migrate to the hypothesis which is producing the slightly better observation measures at the time. This fading of a hypothesis is a common practical problem with particle filter tracking (King & Forsyth, 2000).

In error case (b) in Figure 5, the error is caused by the lateral root consistently having a better measurement model. This error will therefore be present on every run of the algorithm. On inspection, the better measurement appears to be caused by a misrepresentation of the main root in the image. The root here appears very thin. This may be an artefact introduced by the low resolution of the image. However it is caused, the result is that the lateral root provides a higher response to the measurement model and hence the root is traced along this erroneous path.

3.3 Artificial Scenarios

The software was also tested against artificial images. These images were produced using straight lines of a similar colour to the roots. Gaussian noise was applied to the image.

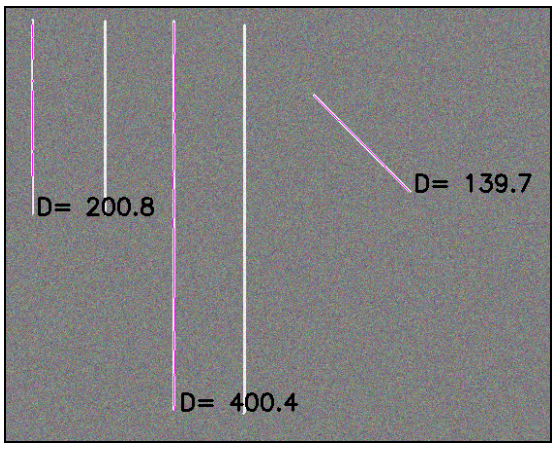


Figure 6: An artificial image with lines of length 200, 400 and 141.4 pixels respectively. Overlaid are example measurements produced by the new software.

The purpose of this experiment was to test the software against a known ground truth measurement. Figure 6 shows one result out of 5 repeats which aimed to test the measuring software against a simple artificial ground truth. The results for the 3 lines measured are presented in Table 3.

Table 3: Results of running the algorithm on artificial data.

Line	True length (pixels)	Average measured by new algorithm (pixels)	Average error (pixels)
1	200	199.9	-0.1
2	400	400.2	0.2
3	141.4	140.3	-1.1

4 SYSTEM EXTENSIONS

The basic system described and tested above has been extended in two ways. First, a method is being developed to automatically detect when the end of the root has been reached. Second, a mixed state particle filter (Isard & Blake 1998) has been incorporated into the framework to allow the labelling of different possible growth modes for the root, such as a gravitropic response. These will be described below.

4.1 Automatic Root Tip Detection

One of the major time consuming and error prone aspects of the root measurement system detailed above is the manual intervention required to stop the line tracing when the end of the root is reached. This was necessary because the premise of the line tracker is that at each iteration the next point on the line definitely does exist somewhere in the image – this assumption is broken when the end of the root is reached. In the absence of a tip detection capability or manual input the tracker will trace whatever produces the best measurement from the image, e.g. see Figure 7 (left).

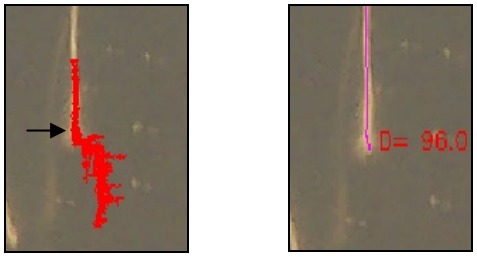


Figure 7: Left: A particle filter tracker will always try and hunt a target even if one does not exist, as the spread of particles past the root tip (indicated by arrow) shows. Right: An example result of the same root image with the new root termination criteria.

The developed method proceeds as follows. During the line tracing phase of the software, the user allows the system to track beyond the end of the root. The graph traversal then proceeds as before, and a final path representing the trace of the root is produced. Now the new step: the measurement probabilities along this path are examined. Figure 8 below shows the trace of log probabilities along a root:

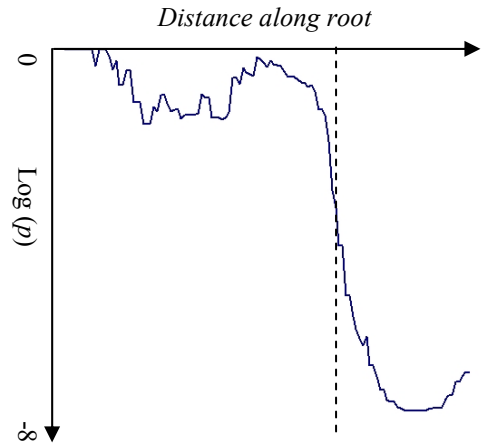


Figure 8: Graph depicting how log of measurement observation probabilities varies along the root. The dashed line marks the approximate end of the root.

Summary statistics of the log probabilities are calculated along the chosen path, and the end of the root is marked as where the current measurement falls below a set number of standard deviations from the mean. This was seen to work well on 7 of the 11 roots in Figure 3 – see figure 7 (right) for an example output.

4.2 Labelling of Growth Modes

It is possible to build into the existing particle filtering framework more than one predictive model to process the particles forward along the root image. This is achieved using a form of mixed state particle filter (Isard & Blake, 1998). Essentially, it is possible to define multiple models for the driving force behind the tracing of the root, and the most appropriate of these will generate higher quality particles at each step. For example, to model gravitropic growth, one model might aim to trace the root left to right across the image, and the second model would aim to trace the root top to bottom. Whichever model prospered the most is naturally selected to label the image – see Figure 9.

Figure 9: Example root trace using a mixed state model consisting of two states, normal growth (white) and gravitropic (black).

5 CONCLUSIONS

5.1 Discussion of Results

The results comparing the software root length measures to the manual measurements show the new technique to produce results to about 2% of the actual measures most of the time. There was a larger error when comparing the new software with the expert user (2.2%) compared to the non-expert (1.7%), however the images in Section 3.2 are more challenging than those in Section 3.1, which may account for some of the increased error also.

Something to be wary of with these kinds of comparisons is using manually marked-up ground truths to compare with the automated measurements. There is an inherent subjectivity in determining the length of the roots, dependant on, for example, the accuracy with which the curves in the roots are traced. The more finely the shape of the root is followed, the longer the measurement. There is similarity here with the coastline measuring problem. Some structures can be thought of as fractal in composition, such as a coastline (Mandelbrot, 1967) or complete root systems (Eshel, 1998). When trying to measure such systems, the scale (or accuracy) with which the waves and perturbations are traced has a bearing on the overall length calculated. This software can be thought of as producing the finest scale estimate of length available at the image resolution, and so is likely to overestimate length compared to a manual measurement. This may be reflected in the results reported in Section 3, with most errors indicating an overestimate of line length.

Even if a user and the new software were to use the same scales of measurement, there is still human error present in the measuring process, which can be quantified by the standard deviation of the manually measured data. The manual measurements in section 3.1 give an average standard deviation of ~2 pixels. Therefore most (99%) of the manually measured lengths can be expected to fall within about 6 pixels (three standard deviations) of the true value for roots of around the length seen in section 3.1. The new software used on these roots has an average relative error of 1.7% which translates to a error of 2.1 pixels on average for these roots, and therefore this software error falls within the expected error bounds of manually entered data.

The time to use the new software was less than the time to take the measurements manually. This should be improved upon still when implementation of the root tip finding algorithm is completed. The system should be less fatiguing to the user as less high-accuracy input is required. This will help to lower the number of mistakes made over the course of measuring many roots.

Labelling of the different growth modes of the root as illustrated in Section 4.2 is also ongoing work, but early results indicate the system can be used for identifying different ways in which a root trace line is produced, as long as trace motion models exist to sufficiently differentiate the modes of production of the line.

5.2 Improving the Reliability

As it stands, the software is still in trial stages and reliability is still being improved. There are a number of possible ways to decrease the number of errors that can occur. One problem is as the particle filter tracks the root towards the tip, it is liable to trace lateral roots if they are long enough and provide a high enough quality measurement, as shown in Figure 5. A simple way to remove this problem is to simply trace the root from the end tip upwards. Due to the geometry of the lateral roots the tracing algorithm is then not presented with viable alternative routes until the lateral roots join and terminate. Therefore, the only way they can be followed is if they lie parallel to the main root for long enough, and are close enough for the particles on the tracing algorithm to ‘jump’ across to the other track. The difficulty with this approach, however, is that the tracker would have to be started on the thinnest, least visible section of the root, which may be hard to detect, and automatic termination of the tracking becomes harder as the delineation at the top of the root is less clear.

Other general improvements include increasing the resolution of the images, as during testing at least some of the mis-tracing of the roots was due to poor representation of the roots in the image. Improving the measurement model may lead to less problems with the system tracking lateral roots. Finally, increasing the number of samples may be beneficial, especially in combination with greater image resolution. However, in such a case speed of

traversal of the graph, which currently is near instantaneous, might become a limiting factor.

5.3 Future Potential of the System

The particle filter approach, with or without mixed state extensions, provides a general framework for matching models of elongated structures to images of those structures. By changing the models used it may be possible to extract descriptions of and measure a wide variety of roots and other plant components. In particular, given higher resolution (e.g. confocal) images showing the cellular structure of the plant, it may be possible to predict (using the motion model) and detect (using the appearance model) higher level structures such as files of cells of similar type.

The ability to recognise state changes by using a mixed state, rather than pure particle filter, also raises the possibility of recognising a wide variety of events during plant growth, of which the onset of gravitropic response may be just the first.

REFERENCES

Abramoff, M. D., Magelhaes, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International, 11,* 36-42.

Bauhus, J. & Messier, C. (1999). Evaluation of fine root length and diameter measurements obtained using RHIZO image analysis. *Agronomy Journal, 91,* 142-147.

Blake, A. & Isard, M. (1998). *Active Contours*. (Second ed.) Springer-Verlag.

Eshel, A. (1998). On the fractal dimensions of a root system. *Plant, Cell and Environment, 21,* 247-251.

Isard, M. & Blake, A. (1998). A mixed-state CONDENSATION tracker with automatic model-switching. *Proc 6th Int.Conf.Computer Vision,* 107-112.

King, O. & Forsyth, D. A. (2000). How does CONDENSATION behave with a finite number of samples? *Proc.ECCV 2000,* 695-709.

Mandelbrot, B. B. (1967). How long is the coast of Britain? *Science, 156,* 636-638.

Pateña, G. F. & Ingram, K. T. (2000). Digital image acquisition and measurement of peanut root minirhizotron images. *Agronomy Journal, 92,* 544.

Qi, X., Qi, J., & Wu, Y. (2007). RootLM: a simple color image analysis program for length measurement of primary roots in *Arabidopsis*. *Plant Root, 1,* 10-16.

Smucker, A. (2007). WR-RIPL 2.0. http://rootimage.css.msu.edu/WR-RIPL/index.html Accessed 20-6-2007.

FULLY-AUTOMATED SEGMENTATION OF TUMOR AREAS IN TISSUE CONFOCAL IMAGES

Comparison between a Custom Unsupervised and a Supervised SVM Approach

Santa Di Cataldo, Elisa Ficarra and Enrico Macii
Dep. of Control and Computer Engineering, Politecnico di Torino, Cso Duca degli Abruzzi 24, 10129, Torino, Italy
santa.dicataldo@polito.it, elisa.ficarra@polito.it, enrico.macii@polito.it

Keywords: Tissue segmentation, tissue confocal images, immunohistochemistry, K-means clustering, Support Vector Machine.

Abstract: In this paper we present a fully-automated method for the detection of tumor areas in immunohistochemical confocal images. The image segmentation provided by the proposed technique allows quantitative protein activity evaluation on the target tumoral tissue disregarding tissue areas that are not affected by the pathology, such as connective tissue. The automated method, that is based on an innovative unsupervised clustering approach, enables more accurate tissue segmentation compared to traditional supervised methods that can be found in literature, such as Support Vector Machine (SVM). Experimental results conducted on a large set of heterogeneous immunohistochemical lung cancer tissue images demonstrate that the proposed approach overcomes the performance of SVM by 8%, achieving on average an accuracy of 90%.

1 INTRODUCTION

Detecting tumor areas in cancer tissue images and disregarding non pathological portions such as connective tissue are critical tasks for the analysis of disease state and dynamics. In fact, by monitoring the activity of proteins involved in the genesis and the development of multi-factorial genetic pathologies we can obtain a useful diagnostic tool. It leads to classify the pathology in a more accurate way through its particular genetic alterations, and to create new opportunities for early diagnosis and personalized predictive therapies (Taneja et al., 2004).

An approach for monitoring and quantifying the protein activity in pathological tissues is to analyze, for example, images of the tissue where the localization of proteins is highlighted by fluorescent marked antibodies that can detect and link the target proteins. The antibodies are marked with particular stains whose intensity is related to protein activity intensity. This procedure is called *immunohistochemistry* (IHC).

The increased use of immunohistochemistry (IHC) in both clinical and basic research settings has led to the development of techniques for acquiring quantitative information from immunostains and automated imaging methods have been developed in an attempt to standardize IHC analysis.

Tissue segmentation for tumor areas detection is the first fundamental step of automated IHC image processing and protein activity evaluation. In fact the quantification of a target protein activity should be performed on tumor portions of the tissue without taking into account the non pathological areas eventually present in the same IHC images. In Figure 1 are reported examples of IHC tissue images where connective tissue (i.e. non tumoral tissue) is outlined in black (for details about these images see Section 2).

Several methods have been proposed in the last few years to perform automated segmentation of tissue images (Demandolx et al., 1997; Nedzved et al., 2000; Malpica et al., 1997; Dybowzki, 2000; Nattkemper 2004). However the most accurate approaches are those that provide a well-suited framework for incorporating primary expert knowledge into the adaptation of algorithms, such as supervised learning algorithm (e.g. Neural Networks, Machine Learning, kernel-based) (Nattkemper 2004). The most prominent algorithm among these is the s*upport vector machine* (SVM) proposed by V.Vapnik (V.Vapnik, 1998) for binary classification. SVM is a theoretically superior machine learning method which has often been shown to achieve great classi-

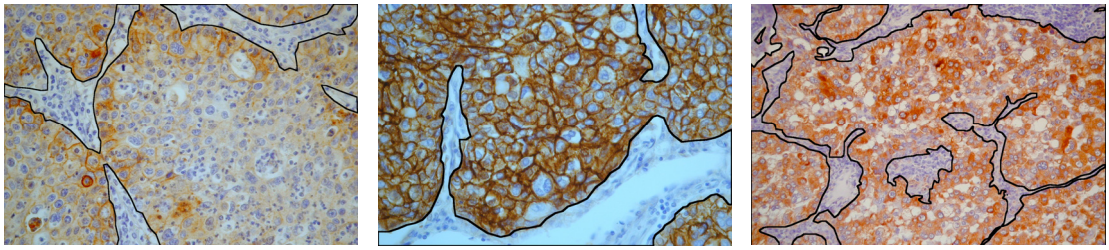


Figure 1: IHC tissue images with connective tissue manually outlined in black (from the left, x400 image with EGF-R positive reactions; x400 image with EGF-R positive reactions; x200 image with TGF-alpha positive reactions).

fication performance compared to other learning algorithms across most application fields and tasks, including image processing and tissue image processing in particular (Angelini et al., 2006; Osuna, 1997). Moreover, the SVM method is more able to handle very high dimensional feature spaces than traditional learning approaches (Muller et al., 2001; Cai, 2003). This is in fact the case of the images targeted by our work.

However, the IHC tissue images we considered in our study present an intrinsic complexity, such as very different characteristics of staining, intensity distribution, considerable variation of tissue shape and/or size and/or orientation and, finally, considerable variation of the signal intensity within the same tissue areas due for example to superimposed staining.

Because of the heterogeneity of the representative features related to each tissue, it is very difficult for the supervised methods to obtain a satisfying fixed classifier able to distinguish between tumor areas (i.e. epithelial tissue) and non cancerous tissue portions (such as connective tissue).

For this reason we designed a fully-automated unsupervised approach that is based only on the characteristics of the input image rather than on a fixed model of the ground truth.

In this paper we present our fully-automated unsupervised method and we compare its performance to that provided by a SVM approach applied on the same IHC tissue image target. We demonstrate that our method enables more accurate tissue segmentation compared with traditional SVM. Experimental results conducted on a large set of heterogeneous immunohistochemical lung cancer images are reported and discussed in Section 4. In Section 2 we detail our fully-automated unsupervised method and we briefly introduce the SVM method. The implementation and the set-up are discussed in Section 3. Finally, the Conclusions are reported in Section 5.

2 METHOD

The images we analyzed in this work were acquired through high-resolution confocal microscopy and show lung cancer tissue cells stained with marked antibodies (see Figure 1). They are characterized by a blue hematoxylin stain as a background colour and a brown DAB stain in cellular regions where a receptor of the EGF-R/*erb-B* or TGF-alpha family is detected (i.e. membranes or cytoplasm, respectively). Cellular nuclei are blue-coloured and show a staining intensity darker than background.

In all the images a remarkable portion of connective or other no cancer tissue components is present, which appears as a blue-coloured mass (since brown DAB-stained cells are only in cancerous tissue) with quite well-defined borders. Connective tissue is usually characterized by shorter inter-cellular distances and smaller nuclei than epithelial component; however, a generalization of this remark is impossible because shape and dimensions distributions of cancer cells are often not predictable. As we outlined in the Introduction, in order to perform accurate and robust cell segmentation and protein activity quantification (Ficarra, 2006) these non cancerous tissue portions have to be identified and isolated from the representative epithelial tissue. Here we present two different segmentation approaches to perform this critical task: i) an unsupervised procedure based on a K-means clustering of brown intensities followed by some morphological and edge-based refinement steps (see Figure 3); ii) a supervised classification of RGB features through Support Vector Machine (see Figure 5).

Experimental results obtained with each approach on the same real-life datasets are presented and compared in Section 4.

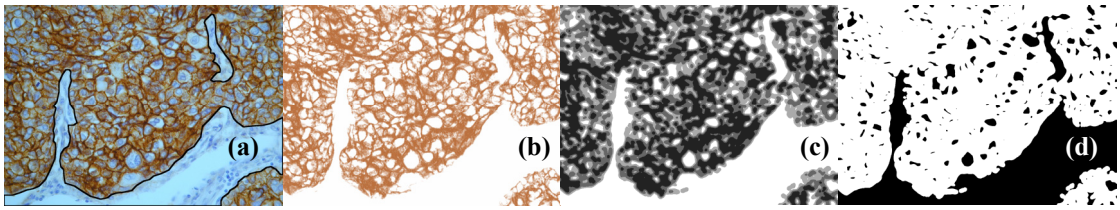


Figure 2: Unsupervised procedure: (a) original IHC image with connective regions manually outlined (in black); (b) pure-DAB image (c) results after K-means clustering (pixels belonging to different clusters are mapped with grey intensity proportional to the cluster centroid); (d) cluster with highest centroid value (in black); as outlined in section 2.1 point 4, some small and round-shaped epithelial particles still have to be removed.

2.1 Unsupervised Procedure

Since non cancerous cells do not show positive reactions at the EGF-R/TGF-alpha receptors, the monochromatic pure-DAB component instead of the original RGB image can be analyzed to perform tissue segmentation: in fact in this simpler color space connective components can be easily identified as wide bright regions with a quite homogeneous appearance (see Figure 2(b)).

An unsupervised learning algorithm (K-means, in our work) can be efficaciously applied to isolate bright regions; then areas which show morphological and edge characteristics which are typical of connective tissue can be selected to refine tissue segmentation.

Figure 3: Unsupervised procedure based on K-means clustering

Main steps of the proposed procedure are (see Figure 3):

1) DAB-Component Separation. To separate pure-DAB from pure-hematoxylin component a color deconvolution algorithm based on stain-specific RGB absorption is applied on the original RGB image (Ruifrok 2001, 2004); differently from classical color segmentation approaches based on transformation of RGB information to HSI or to another specific color representation (Brey, 2003), this method has been demonstrated to perform a good color separation even with colocalized stains. This critical condition, due to chemical reactions of stains linking the target proteins and to the tissue superposition during the slicing of samples before image acquisition, is very common in the images targeted by our method.

For this step, the free color deconvolution plugin developed by G. Landini was integrated to our algorithm.

2) Preprocessing. In pure-DAB images, connective tissue can be differentiated from epithelial tissue through its higher intensity (see Figure 2(b)); anyway some preprocessing is needed in order to homogenize and separate the intensity distributions of the two tissues, thus improving K-means' performance.

First of all, a mean filter is performed: this operation replaces each pixel value with the average value in its neighbourhood, thus smoothing intensity peaks and decreasing the influence of single non-representative pixels. Then a minimum filter is applied. The filter replaces pixels values with the minimum intensity values in their neighbourhood: this transformation reduces the intensity dynamic and performs a further separation of connective and epithelial intensity distributions, since the former shows minimum values higher than the latter.

3) K-Means Clustering. To isolate bright pixels belonging to connective tissue a K-means clustering, the well-known unsupervised learning algorithm (Jain, 1988) which iteratively partitions a given dataset into a fixed number of clusters, is applied; this iterative partitioning minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances. Thus the procedure minimizes the so-called *objective function*, J in Equation 1, where k is the number of clusters, n is the number of data points and the quadratic expression is the distance measure between a data point $x_i^{(j)}$ and the current cluster centroid c_j.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \quad (1)$$

The cluster with the highest centroid value is selected as representative of the connective tissue (see

Figure 2(c)). The number of clusters k was empirically set to four (see Section 3.1 for details about the parameter set-up).

4) Refinement by Size and Circularity Analysis. Bright epithelial regions with low EGF-R/TGF-alpha activity have to be removed from the connective cluster to refine tissue segmentation. As shown in Figure 2(d), a large number of these regions are approximately round-shaped and are considerably smaller than connective mass: then a selective removal of particles with a low area and a high circularity compared to threshold values T_S and T_C is performed (parameters set-up in Section 3.1).

Equation 2 shows the proposed index for circularity evaluation (a value of 1 indicates a perfect circle, a value approaching 0 an increasingly elongated polygon).

$$Circularity = 4\pi \frac{Area}{Perimeter^2} \quad (2)$$

5) Refinement by Gradient Magnitude Analysis. Other bright epithelial regions can be removed from the connective cluster through their edge characteristics, since connective tissue usually shows a well-defined boundary w.r.t. epithelial background in terms of intensity gradient variation. On the base of this remark, in this step areas which show along their boundary a percentage of edge pixels (i.e. pixels with high gradient intensity variation w.r.t. background) lower than a threshold value T_E are selectively removed from connective cluster (parameter set-up in Section 3.1). Edge detection is performed through a Sobel detector followed by automated intensity global thresholding.

2.2 Supervised Procedure

An alternate approach for tissue segmentation is supervised learning; for this purpose a *Support Vector Machine* (SVM) classification is proposed.

The SVM (Vapnik, 1998) is a theoretically superior machine learning method which has often been shown to achieve great classification performance compared to other learning algorithms across most application fields and tasks including image processing (Statnikov, 2005).

Here we propose a procedure based on binary SVM classification, in which the input elements (in this work, small tissue regions) are associated to one of two different classes, connective or epithelial, on the base of a set of representative characteristics, the *features vector*. To perform a reliable classification, the SVM is previously trained with a set of elements whose class is well-known, the so-called *training instances*.

The classification is based on the implicit mapping of data to a higher dimensional space via a *kernel function* and on the consequent solving of an optimization problem to identify the *maximum-margin hyperplane* that separates the given training instances (see Figure 4).

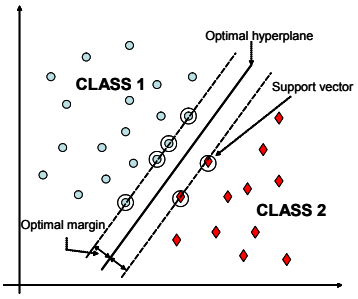


Figure 4: Maximum-margin hyperplane in SVMs (linearly separable case). The boundary training instances (*support vectors*) are indicated by an extra circle.

This hyperplane is calculated on the base of boundary training instances (i.e. elements with characteristics which are border-line between the two classes), the so-called *support vectors*; new instances are then classified according to the side of the hyperplane they fall into.

In order to handle linearly nonseparable data, the optimization *cost function* includes an error minimization term to penalize the wrongly classified training instances.

See the references provided in the text for a technical description of SVMs.

Figure 5: Supervised procedure based on SVM.

Our proposed supervised procedure for tissue segmentation consists in three main steps (see Figure 5):

1) Training Features Extraction. In order to obtain a good generalization of the SVM, a skilled operator was asked to select from a large number of real-life tissue images small rectangular regions wherein both connective and epithelial tissue were present. The images showed various staining levels and very different characteristics of tissue shape and intensity distribution.

In each representative sample the operator manually traced the boundaries of connective and epithelial tissue. Then a NxN square sliding window was

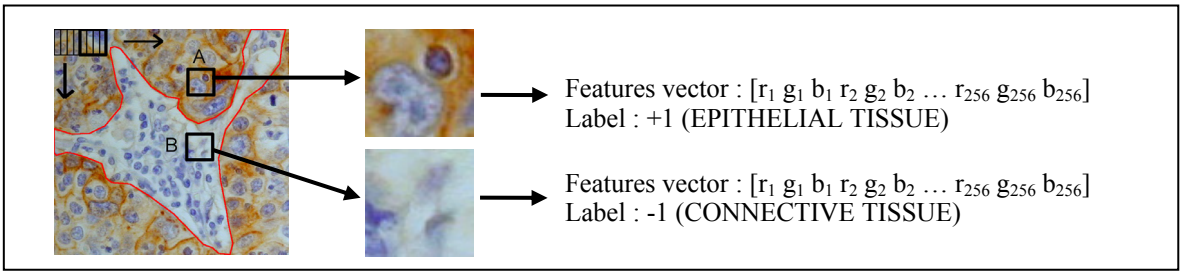


Figure 6: Generation of the features vectors for SVM training. A NxN square window is horizontally and vertically shifted on the sample, thus covering the entire surface of the image. For each shift a features vector is generated with RGB values of 256 equally spaced pixels, as for Window A and Window B. Epithelial instances are labelled with a +1, connective instances with a -1.

horizontally and vertically shifted over the samples (shift value *s*), thus covering the entire surface of the image; for each shifted window, a features vector was generated with the RGB values of 256 equally-spaced pixels (see Figure 6, parameters set-up in Section 3.2).

In this way, a features vector of 3x256 variables was created for each single shift.

A +1 label was assigned to windows with a prevalence of epithelial tissue pixels, a -1 label to windows with a prevalence of connective tissue pixels.

2) Training. The labelled features vectors were fed into the SVM for the training; for details about the parameters set-up see Section 3.2.

3) Classification. The optimized SVM obtained in the training step is used to perform tissue classification for new images.

For this purpose, the input images are processed to generate features vectors as in step 1 which are fed into the trained SVM. At the end of the classification, the SVM automatically associates positive labels to epithelial patterns and negative labels to connective patterns. The output is then processed to reconstruct a two-dimensional result as in Figure 8.

3 IMPLEMENTATION

The algorithm was implemented in Java as a plugin for ImageJ, a public domain image analysis and processing software which runs on all the standard operating systems (Windows, Mac OS, Mac OS X and Linux): therefore it is totally hardware-independent, flexible and upgradeable. We inherited the whole class hierarchy of the open-source *ImageJ 1.37 API* and the free plugins for color deconvolution (Landini) and K-means clustering (Sacha) and we implemented our own functions and classes. A user-friendly interface enables the user to set different parameters values without modifying the source code.

For the supervised procedure we used the *cSVM* tool for binary classification (Anguita, 2005), since it uses the state-of-art optimization method SMO, i.e. *Sequential Minimal Optimization* (Platt, 1999). This cSVM tool implements the algorithm described in (Wang, 2004), which was successfully used to solve different real world problems. Our ImageJ plugins for features vectors generation and output reconstruction were integrated to the SVM tool.

The parameters of the proposed algorithms were empirically tuned by a skilled operator after running several experiments on a large dataset of real tissue images which showed very different characteristics of staining intensity, resolution, EGF-R/TGF-alpha activity level, tissue shape. In the following subsections, we report some details about the implementation of both the unsupervised and the supervised classification procedures and we outline the experimental set-up of the main parameters.

3.1 Unsupervised Procedure

The **number of clusters** *k* (see Section 2.1 point 3) was set to 4 after running the algorithm with values varying from 2 to 5 and evaluating each time K-means performance in terms of sensibility (power to detect connective components) and selectivity (power to avoid misclassification of epithelial components). For values lower than 4 we often experienced a very good sensibility but a not sufficient selectivity; for higher values the sensibility was frequently poor. A *k* value equal to 4 assured a good performance of K-means in all the tested images.

The **size threshold** T_S (see Section 2.1 point 4) was varied from 1000 to 5000 pixels with a step of 1000 and was finally set to 3000. Increasing values led to a progressive improvement of selectivity in the connective tissue selection; with values higher

Table 1: Experimental results of unsupervised and supervised classifications. As outlined in Section 3.2, in supervised classification two different SVMs trained respectively with x200 and x400 samples were used (the number of training instances extracted from each dataset and the total number of training instances are reported for both x200 and x400 classifiers). Training instances were removed from the validation dataset, which was considerably larger.

		UNSUPERVISED ALGORITHM			***SUPERVISED ALGORITHM***					
Dataset		***Sensibility (%)***	***Selectivity (%)***	***Accuracy (%)***		*Number of training instances*	*Number of validation instances*	***Sensibility (%)***	***Selectivity (%)***	***Accuracy (%)***
x200	1	81,89	90,54	84,77		1692	28308	57,91	91,38	69,07
	2	94,64	84,94	91,41		912	20263	94,05	79,20	89,10
	3	95,21	97,99	96,14		220	20192	91,09	94,75	92,31
	4	86,60	87,32	86,84		408	19142	84,41	91,18	86,66
					tot	3232	91137			
x400	5	91,77	86,20	89,91		558	6942	67,48	82,35	72,43
	6	91,30	78,56	87,05		640	6860	66,48	90,02	74,32
	7	99,67	93,33	97,56		252	7248	93,53	87,46	91,51
	8	89,21	86,28	88,23		300	5888	87,29	85,39	86,66
					tot	1750	28688			

than 3000 the lack in sensibility was often not acceptable. Similarly, the **circularity threshold** T_C (see Section 2.1 point 4) was decreased from 0,9 to 0,3. A value of 0,7 assured a good selectivity enhancement without altering sensibility in any of the images.

The **edge threshold** T_E (see Section 2.1 point 5) was increased from 20% to 35% with a step of 5%, evaluating each time the parameter performance in terms of selectivity enhancement and sensibility preservation. A value of 25% assured the best improvement in selectivity without altering sensibility in any of the tested images.

3.2 Supervised Procedure

The **window size** N for features vectors generation (see Section 2.2 point 1) should grant a visible differentiation between connective and epithelial tissue; since nuclei are blue-colored and quite similar in both the tissues, the window has to be large enough to contain a whole nucleus and some surrounding tissue. On the other hand, lower-sized windows allows a better selectivity.

After running several experiments with values varying from 16 to 72 pixels, N was set to 32 for x200 images and to 64 for x400 images.

Since the optimal window size depends on image resolution, x200 and x400 images were respectively classified with SVM trained with x200 and x400 samples.

The **shift** value s (see Section 2.2 point 1) was set to $N/4$, which granted the best compromise between selectivity of classification and computational time.

After running experiments with linear, gaussian and polynomial kernels, we finally chose the normalized **polynomial kernel** shown in Equation 3, where x_1 and x_2 are feature vectors, n=768 is the input space dimension and p=2 is the kernel hyperparameter; see (Wang, 2004) for technical details).

$$\boldsymbol{K}(x_1, x_2) = \frac{(x_1 \cdot x_2 + n)^p}{\sqrt{(x_1 \cdot x_1 + n)^p}\sqrt{(x_2 \cdot x_2 + n)^p}} \quad (3)$$

4 EXPERIMENTAL RESULTS

We tested the performance of both the algorithms on a large dataset extracted from real tissue images which presented positive reactions at the EGF-R or at the TGF-alpha receptor activation (see Figure 1 for examples); reactions are localized in cellular membranes for EGF-R and in cytoplasm for TGF-alpha. Images were acquired from different samples with two different enlargements, x200 or x400.

A skilled operator was asked to manually draw the boundaries of connective tissue in each of the testing datasets. The manual segmentations performed by the operator were pixel-by-pixel compared to those obtained by both the unsupervised and the supervised algorithms. Connective tissue selection was evaluated in terms of *sensibility* (i.e. power to detect connective tissue) and *selectivity* (power to avoid misclassification of non-connective tissue): for this purpose, the percentage of respectively connective and non-connective pixels which were equally classified by manual and automated segmentation was calculated. The segmentation *accuracy* was then calculated as weighted average of sensibility and selectivity, as in Equation 4.

$$\boldsymbol{Accuracy} = \frac{2}{3} \cdot \boldsymbol{Sensibility} + \frac{1}{3} \cdot \boldsymbol{Selectivity} \quad (4)$$

Different weights were used because sensibility is more critical for automated measures of protein

activity, which is the principal application targeted by our method: in fact, in order to obtain a reliable measure, it is fundamental to eliminate as much as possible non representative tissues from the range of interest; on the contrary, erroneous removal of some epithelial regions is more tolerable, since it has a lower influence on the final measure.

Results obtained for both the automated algorithms are reported in Table 1. The number of training instances extracted from each dataset and the total number of training instances are reported too for both x200 and x400 SVMs. The classification performance was evaluated on a large validation dataset which did not include the patterns used for training.

Some examples of tissue segmentation are shown in Figure 8.

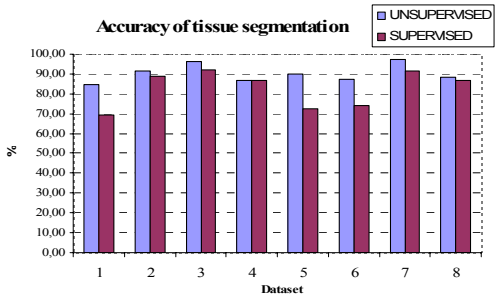


Figure 7: Accuracy of tissue segmentation; comparison between unsupervised and supervised procedure.

As shown in Table 1 and Figure 7, our unsupervised procedure achieved the best results: this method performed tissue segmentations highly comparable with those provided by the skilled operator in all the testing datasets; mean accuracy was 90,24%, with values generally around 90% and always above approximately 85%. SVM performed worse in all the tested datasets; mean accuracy was about 7,5% lower than our unsupervised method.

As we previously outlined, SVM is a theoretically superior machine learning method which has often been shown to achieve great classification performance compared to other learning algorithms across most application fields and tasks including image processing (Angelini et al., 2006; Cai, 2003; Muller et al., 2001; Osuna, 1997). However, in this case its classification performance was poor because of the intrinsic complexity of the images targeted by our method: in fact, these images showed very different characteristics of staining, tissue shape and intensity distribution. Because of the heterogeneity of the representative features of each class, it was impossible for the supervised method to obtain a satisfying separability of connective and epithelial tissue.

Images heterogeneity was less critical for the unsupervised approach, since differently from SVMs it is based only on the characteristics of the input image and not on a fixed model of the ground truth.

On the other hand, our unsupervised method's selectivity is influenced by tissue composition: in fact, since the number of clusters is a-priori fixed, some epithelial regions with low brown staining are often misclassified in images without any connective tissue.

Despite this eventuality is unlikely, since pure-epithelial tissue samples are very uncommon (and we reasonably suppose that the operator would escape the automated tissue segmentation in this case), we are working on the solution of the problem: in particular, the introduction of an adaptive number of clusters is in development.

As regards the supervised approach, other learning methods such as neural networks and artificial neural networks (ANN) will be tested in the future.

5 CONCLUSIONS

We presented a fully-automated unsupervised tissue image segmentation method that allows to distinguish tumor areas in immunohistochemical images and disregard non pathological areas such as connective tissue. This procedure is critical for automated protein activity quantification in tumor tissues in

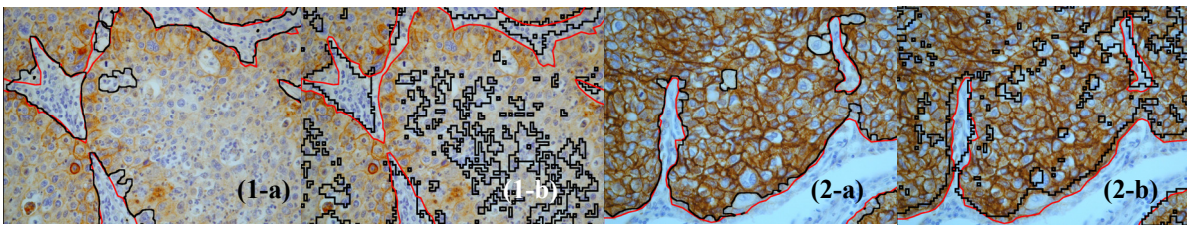


Figure 8: Examples of tissue segmentation performed by the unsupervised (a) and the supervised (b) algorithm (manual segmentation in red, automated segmentation in black).

order to analyze the pathology dynamics and development.

We described the original processing steps we designed. Finally, we carried out an extensive experimental evaluation on a large set of heterogeneous images that demonstrated the high accuracy achievable by the proposed technique (90% on average) compared to a more traditional approach based on Support Vector Machines (SVM).

As future work, we will compare the proposed approach to artificial neural networks (ANN), and we will eventually study the possibility of their integration.

ACKNOWLEDGEMENTS

We acknowledge the Dep. of Pathology of the S.Luigi Hospital of Orbassano in Turin, Italy, for providing IHC images and for the helpful and stimulating discussions.

REFERENCES

Angelini, E, Campanini, R., Iampieri, E., Lanconelli, N. Masotti, M., Roffilli, M., 2006. Testing the performances of different image representation for mass classification in digital mammograms. *Int. J. Mod. Phys*. 17(1):113-131.

Anguita, D., Boni, A., Ridella, S., Rivieccio F., Sterpi, D., 2005. Theoretical and Practical Model Selection Methods for Support Vector Classifiers. *Springer, Studies in Fuzziness and Soft Computing, Support Vector Machines: Theory and Application.*

Brey, E.M., Lalani, Z., Hohnston, C., Wong, M., McIntire, L.V., Duke, P.J., Patrick, C.W., 2003. Automated selection of DAB-labeled tissue for immunohistochemical quantification. In *J. Histochem. Cytochem.,* 51(5), pp.575-584.

Cai, C.Z., W.L.Wang, Y.Z: Chen, 2003. *Int.J.Mod.Phys*. 14:575.

Demandolx D, Davoust J. , 1997. Multiparameter image cytometry: from confocal micrographs to subcellular fluorograms. *Bioimaging*. 4:159-169.

Dybowski R., 2000. Neural computation in medicine: perspectives and prospects. *Proc. ANNIMAB-1*. pp. 27-36.

E. Osuna, R. Freund, F. Girrosi, 1997. Training Support Vector Machines: an Application to Face Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*. pp. 130.

Ficarra, E., Macii, E., De Micheli, G., 2006. Computer-aided evaluation of protein expression in pathological tissue images. In *Proc. of IEEE CBMS'06.*, pp.413-418.

Jain, A.K., Dubes, R.C., 1988. *Algorithms for clustering data,* Prentice Hall.

Landini, G., 2007. Software, http://www.dentistry.bham.ac.uk/landinig/software/software.html

Malpica N, de Solorzano CO, Vaquero JJ, Santos A, Vallcorba I, Garcia-Sagredo JM, del Pozo F, 1997. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*. 28(4): 289-297.

Muller K.R:, S. Mika,G. Ratsch, K. Tsuda, 2001. *IEEE Trans. Neural Networks*, 12:181.

Nattkemper, T.W., 2004. Automatic segmentation of digital micrographs: A survey. *Medinfo*, 11(Pt 2):847-51.

Nedzved A, Ablameyko S, Pitas I., 2000. Morphological segmentation of histology cell images. *ICPR*. 1:500-3.

Platt, J., 1999. Fast training of support vector machines using sequential minimal optimization. In Scholkopf, B., *Advances in kernel methods-support vector learning*. MIT Press, Cambridge, MA, USA.

Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, rsb.info.nih.gov/ij/.

Ruifrok, A.C., Johnston, D.A., 2001. Quantification of histochemical staining by color deconvolution. In *Anal.Quant.Cytol.Histol.,* 23(4), pp.291-299.

Ruifrok, A.C., Katz, R., Johnston, D., 2004. Comparison of quantification of histochemical staining by Hue-Saturation-Intensity (HSI) transformation and color deconvolution. In *Appl. Immunohisto. M. M.,* 11(1), pp.85-91.

Sacha, J., K-means clustering, http://ij-plugins.sourceforge.net/plugins/clustering/index.html.

Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D., Levy, S., 2005. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. In *Bioinformatics,* 21(5), pp.631-643.

Taneja, T.K., SK.Sharma Markers of small cell lung cancer. *World Journal of Surgical Oncology*, Vol(2):10.

Vapnik, V., 1998. *Statistical learning theory,* Wiley-Interscience, New York, NY, USA.

Wang, L., 2004. *Support vector machines: theory and applications,* Springer.

MULTI-CHANNEL BIOSIGNAL ANALYSIS FOR AUTOMATIC EMOTION RECOGNITION

Jonghwa Kim and Elisabeth André
Institute of Computer Science, University of Augsburg, Eichleitnerstr. 30, D-86159 Augsburg, Germany
kim@ieee.org, andre@informatik.uni-augsburg.de

Keywords: Biosignal, emotion recognition, physiological measures, skin conductance, electrocardiogram, electromyogram, respiration, affective computing, human-computer interaction, musical emotion, autonomic nervous system, arousal, valence, feature extraction, pattern recognition.

Abstract: This paper investigates the potential of physiological signals as a reliable channel for automatic recognition of user's emotial state. For the emotion recognition, little attention has been paid so far to physiological signals compared to audio-visual emotion channels such as facial expression or speech. All essential stages of automatic recognition system using biosignals are discussed, from recording physiological dataset up to feature-based multiclass classification. Four-channel biosensors are used to measure electromyogram, electrocardiogram, skin conductivity and respiration changes. A wide range of physiological features from various analysis domains, including time/frequency, entropy, geometric analysis, subband spectra, multiscale entropy, etc., is proposed in order to search the best emotion-relevant features and to correlate them with emotional states. The best features extracted are specified in detail and their effectiveness is proven by emotion recognition results.

1 INTRODUCTION

In human communication, expression and understanding of emotions facilitate to complete the mutual sympathy. To approach it in human-machine interaction, we need to equip machines with the means to interpret and understand human emotions without input of user's translated intention. Hence, one of the most important prerequisites to realize such an advanced user interface is a reliable emotion recognition system which guarantees acceptable recognition accuracy, robustness against any artifacts, and adaptability to practical applications. It is about to model, analyze, process, train, and classify emotional features measured from the implicit emotion channels of human communication, such as speech, facial expression, gesture, pose, physiological responses, etc. In this paper we concentrate on finding emotional cues from various physiological measures.

Recently many works on engineering approaches to automatic emotion recognition have been reported. For an overview we refer to (Cowie et al., 2001). Particularly, most efforts have been taken to recognize human emotions using audiovisual channels of emotion expression, facial expression, speech, and gesture. Relatively little attention, however, has been paid so far to using physiological measures. Reasons are some significant limitations resulting from the use of physiological signals for emotion recognition. The main difficulty lies in the fact that it is a very hard task to uniquely map subtle physiological patterns onto specific emotional states. As an emotion is a function of time, context, space, culture, and person, physiological patterns may also widely differ from user to user and from situation to situation.

In this paper, we treat all essential stages of automatic emotion recognition system using physiological measures, from data collection up to classification of four typical emotions (joy, anger, sadness, pleasure) using four-channel biosignals. The work in this paper is novel in trying to recognize naturally induced musical emotions using physiological changes, in acquiring the physiological dataset through everyday life recording over many weeks from multiple subjects, in finding emotion-relevant ANS (autonomic nervous system) specificity through various feature contents, and in designing an emotion-specific classification method. After the calculation of a great number of features (a total of 110 features) from various feature domains, we try to identify emotion-relevant features using the backward feature selection method combined with a linear classifier. These features can be directly used to design affective human-machine interfaces for practical applications.

2 RELATED RESEARCH

A significant amount of work has been conducted by Picard and colleagues at MIT Lab who showed that certain affective states may be recognized by using physiological data including heart rate, skin conductivity, temperature, muscle activity and respiration velocity (Healey and Picard, 1998). They used personalized imagery to elicit target emotions from a single subject who had two years' experience in acting, and achieved overall 81% recognition accuracy in eight emotions by using hybrid linear discriminant classification (Picard et al., 2001). Nasoz et al. (Nasoz et al., 2003) used movie clips based on the study by Gross and Levenson (Gross and Levenson, 1995) for eliciting target emotions from 29 subjects and achieved best emotion classification accuracy of 83% through the Marquardt Backpropagation algorithm (MBP). More recently, interesting user-independent emotion recognition system is reported by Kim et al. (Kim et al., 2004). They developed a set of recording protocols using multimodal stimuli (audio, visual, and cognitive) to evoke targeted emotions (sadness, stress, anger, and surprise) from the 175 children aged from five to eight years. Classification ratio of 78.43% for three emotions (sadness, stress, and anger) and 61.76% for four emotions (sadness, stress, anger, and surprise) has been achieved by adopting support vector machine as pattern classifier.

Note that the recognition rates in the privious works should be strongly dependent on the datasets they used and context of subjects. Moreover, the physiological datasets used in most of the previous works are gathered by using visual elicitation materials in a lab setting. The subjects then "tried and felt" or "acted out" the target emotions while looking at selected photos or watching movie clips that are carefully prearranged to the emotions. It means, extremely speaking, that the recognition results were achieved for specific users in specific contexts with the "forced" emotional states.

3 MUSICAL EMOTION INDUCTION

A well established mechanism of emotion induction would be either to imagine or to recall from individual memory. Emotional reaction can be triggered by a specific cue and be evoked by an experimental instruction to imagine certain events. On the other hand, it can spontaneously be resurged in memory. Music is a pervasive element accompanying many highly significant events in human social life and particular pieces of music are often connected to significant personal memories. This claims that certain music can be a powerful cue in bringing emotional experiences from memory back into awareness. Since music listening is often done by an individual in isolation, the possible artifacts by social masking and social interaction may be minimized in the experiment. Furthermore, like odors, music may be treated at lower levels of the brain that are particularly resistant to modifications by later input, contrary to cortically based episodic memory (LeDoux, 1992).

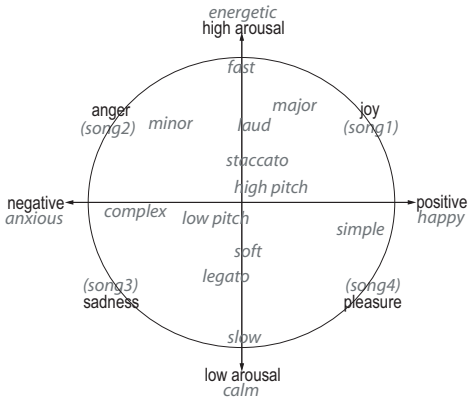


Figure 1: Reference emotional cues in music based on the 2D emotion model. Metaphoric cues for song selection: song1 (enjoyable, harmonic, dynamic, moving), song2 (noisy, loud, irritating, discord), song3 (melancholic, sad memory), song4 (blissful, pleasurable, slumberous, tender).

To collect a database of physiological signals in which the targeted emotions (joy, anger, sadness, pleasure)[1] can be *naturally* reflected without any deliberate expression, we decided to use musical induction method, recording physiological signals while the subjects are listening to different music songs. The subjects were three males (two students and an academic employee) between 25-38 years old and enjoyed listening to music everyday. The subjects individually handpicked four music songs by themselves that should spontaneously evoke their emotional memories and certain moods corresponding to the four target emotions. Figure 1 shows the musical emotion model referred to for the selection of their songs. Generally, emotional responses to music would vary greatly from individual to individual depending on their unique past experiences. Moreover,

[1]We note that these four expression words are used to cover each quadrant in the 2D emotion model, i.e. joy should represent *all* emotions with high arousal and positive valence, anger with high arousal and negative, sadness with low arousal and negative, and pleasure with low arousal and positive valence.

cross-cultural comparisons in literature suggest that the emotional responses can be quite differentially emphasized by different musical cultures and training. This is why we advised the subjects to choose themselves the songs that recall their individual special memories with respect to the target emotions.

For the experiment, we prepared a quiet listening room in our institute in order to ensure the subjects to unaffectedly feel the emotions from the music. For the recording, the subject needs to position himself the sensors by instruction posters in the room, to apply the headphones, and to select a song from his song list saved in the computer. When he does mouse-click just at the start of recording, the recording and music systems are automatically setting up by preset values for each song. Recording schedule was determined by the subjects themselves too, at any time when they will listen to music and which song they choose. It means, different from methods used in other studies, that the subjects were not forced to participate in a lab setting scenario and to use prespecified stimulation materials. We believe that this voluntary participation of the subjects during our experiment might be a help to obtain a high-quality dataset with natural emotions. The physiological signals are acquired us-

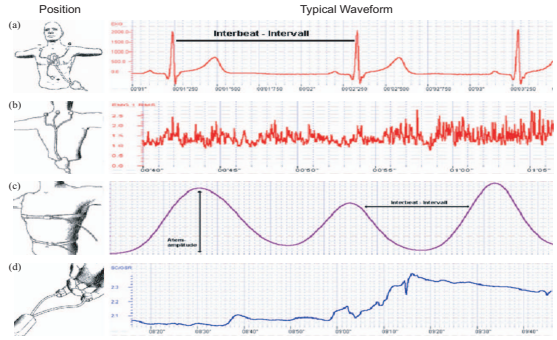


Figure 2: Position and typical waveforms of the biosensors: (a) ECG, (b) EMG, (c) RSP, (d) SC.

ing the Procomp Infiniti™ (www.mindmedia.nl) with four biosensors, electromyogram (EMG), skin conductivity (SC), electrocardiogram (ECG), and respiration (RSP). The sampling rates are 32 Hz for EMG, SC, and RSP, and 256 Hz for ECG. The positions and typical waveforms of the biosensors we used are illustrated in Fig. 2. We used pre-gelled single Ag/AgCl electrodes for ECG and EMG sensors and standard single Ag/AgCl electrodes fixed with two finger bands for SC sensor. A stretch sensor using latex rubber band fixed with velcro respiration belt is used to capture breathing activity of the subjects. It can be worn either thoracically or abdominally, over clothing.

During the three months, a total of 360 samples (90 samples for each emotion) from three subjects is collected. Signal length of each sample is between 3-5 minutes depending on the duration of the songs.

4 METHODOLOGY

Overall structure of our recognition system is illustrated in Figure 3.

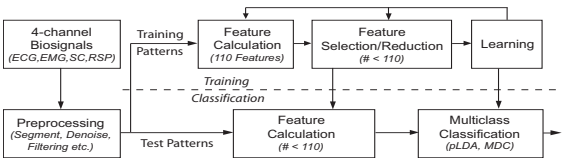


Figure 3: Block diagram of supervised statistical classification system for emotion recognition.

4.1 Preprocessing

Different types of artifacts were observed in all the four channel signals, such as transient noise due to movement of the subjects during the recording, mostly at the begin and end of the each recording. Thus, uniformly for all subjects and channels, we segmented the signals into final samples with fixed length of 160 seconds by cutting out from the middle part of each signal. Particularly to the EMG signal, we needed to pay closer attention because the signal contains artifacts generated by respiration and heart beat (Fig. 4). It was due to the position of EMG sensor at the nape of the neck. For other signals we used pertinent lowpass filters to remove the artifacts without loss of information.

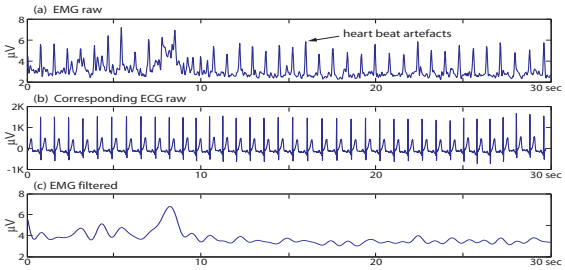


Figure 4: Example of EMG signal with heart beat artifacts and denoised signal.

4.2 Measured Features

From the four channel signals we calculated a total of 110 features from various analysis domains including conventional statistics in time series, frequency

domain, geometric analysis, multiscale sample entropy, subband spectra, etc. For the signals with non-periodic characteristics, such as EMG and SC, we focused on capturing the amplitude variance and localizing the occurrences (number of transient changes) in the signals.

4.2.1 Electrocardiogram (ECG)

To obtain subband spectrum of the ECG signal we used the typical 1024 points fast Fourier transform (FFT) and partitioned the coefficients within the frequency range 0-10 Hz into eight non-overlapping subbands with equal bandwidth. First, as features, power mean values of each subband and fundamental frequency (F0) are calculated by finding maximum magnitude in the spectrum within the range 0-3 Hz. To capture peaks and their locations in subbands, subband spectral entropy (SSE) is computed for each subband. To compute the SSE, it is necessary to convert each spectrum into a probability mass function (PMF) like form. Eq. 1 is used for the normalization of the spectrum.

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i}, \qquad \text{for } i = 1 \ldots N \tag{1}$$

where X_i is the energy of i^th frequency component of the spectrum and $\tilde{\mathbf{x}} = \{x_1 \ldots x_N\}$ is to be considered as the PMF of the spectrum. In each subband the SSE is computed from $\tilde{\mathbf{x}}$ by

$$H_{\text{sub}} = -\sum_{i=1}^{N} x_i \cdot \log_2 x_i \tag{2}$$

By packing the eight subbands into two bands, i.e., subbands 1-3 as low frequency (LF) band and subbands 4-8 as high frequency (HF) band, the ratios of LF/HF bands are calculated from the power mean values and the SSEs.

To obtain the HRV (heart rate variability) from the continuous ECG signal, each QRS complex is detected and the RR intervals (all intervals between adjacent R waves) or the normal-to-normal (NN) intervals (all intervals between adjacent QRS complexes resulting from sinus node depolarization) are determined. We used the QRS detection algorithm of Pan and Tompkins (Pan and Tompkins, 1985) in order to obtain the HRV time series. Figure 5 shows example of R wave detection and interpolated HRV time series, referring to the increases and decreases over time in the NN intervals.

In time-domain of the HRV, we calculated statistical features including mean value, standard deviation of all NN intervals (SDNN), standard deviation of first difference of the HRV, the number of pairs

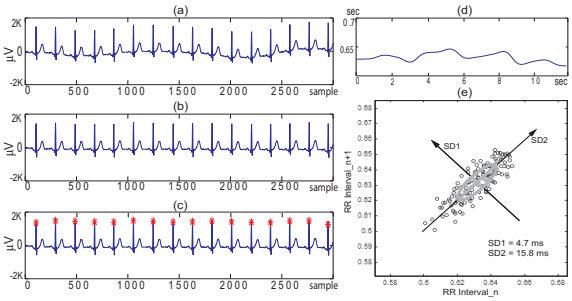


Figure 5: Example of ECG Analysis: (a) raw ECG signal with respiration artifacts, (b) detrended signal, (c) detected RR interbeats, (d) interpolated HRV time series, (e) Poincaré plot of the HRV time series.

of successive NN intervals differing by greater than 50 ms (NN50), the proportion derived by dividing NN50 by the total number of NN intervals. By calculating the standard deviations in different distances of RR interbeats, we also added Poincaré geometry in the feature set to capture the nature of interbeat (RR) interval fluctuations. Poincaré plot geometry is a graph of each RR interval plotted against the next interval and provides quantitative information of the heart activity by calculating the standard deviations of the distances of the $R-R(i)$ to the lines $y = x$ and $y = -x + 2 * R - R_m$, where $R - R_m$ is the mean of all $R-R(i)$, (Kamen et al., 1996). Figure 5.(e) shows an example plot of the Poincaré geometry. The standard deviations SD_1 and SD_2 refer to the fast beat-to-beat variability and longer-term variability of $R-R(i)$ respectively.

Entropy-based features from the HRV time series are also considered. Based on the so-called *approximate entropy* and *sample entropy* proposed in (Richmann and Moorman, 2000), a multiscale sample entropy (MSE) has been introduced (Costa et al., 2005) and successfully applied to physiological data, especially for analysis of short and noisy biosignal. Given a time series $\{X_i\} = \{x_1, x_2, ..., x_N\}$ of length N, the number ($n_i^{(m)}$) of similar m-dimensional vectors $y^{(m)}(j)$ for each sequence vectors $y^{(m)}(i) = \{x_i, x_{i+1}, ..., x_{i+m-1}\}$ is determined by measuring their respective distances. The relative frequency to find the vector $y^{(m)}(j)$ within a tolerance level δ is defined by

$$C_i^{(m)}(\delta) = \frac{n_i^{(m)}}{N - m + 1} \tag{3}$$

The approximate entropy, $h_A(\delta, m)$, and the sample entropy, $h_S(\delta, m)$ are defined as

$$h_A(\delta,m) = \lim_{N\to\infty}[H_N^{[m]}(\delta) - H_N^{(m+1)}(\delta)], \quad (4)$$

$$h_S(\delta,m) = \lim_{N\to\infty} -\ln\frac{C^{(m+1)}(\delta)}{C^{(m)}(\delta)}, \quad (5)$$

where

$$H_N^{(m)}(\delta) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln C_i^{(m)}(\delta), \quad (6)$$

Because of advantage of being less dependent on time series length N, we applied the sample entropy h_S to coarse-grained versions ($y_j^{(\tau)}$) of the original HRV time series $\{X_i\}$,

$$y_j(\tau) = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le N/\tau, \quad \tau = 1,2,3,\ldots \quad (7)$$

The time series $\{X_i\}$ is first divided into N/τ segments by non-overlapped windowing with length of scale factor τ and then the mean value of each segment is calculated. Note that for scale one $y_j(1) = x_j$. From the scaled time series $y_j(\tau)$ we obtain the m-dimensional sequence vectors $y^{(m)}(i,\tau)$. Finally, we calculate the sample entropy h_S for each sequence vector $y_j(\tau)$. In our analysis we used $m = 2$ and fixed $\delta = 0.2\sigma$ for all scales, where σ is the standard deviation of the original time series x_i. Note that using the fixed tolerance level δ as a percentage of the standard deviation corresponds to initial normalizing of the time series and it thus enables that h_S does not depend on the variance of the original time series, but only on their sequential ordering.

In frequency-domain of the HRV time series, three frequency bands are of interest in general; very-low frequency (VLF) band (0.003-0.04 Hz), low frequency (LF) band (0.04-0.15 Hz), and high frequency (HF) band (0.15-0.4 Hz). From these subband spectra, we computed dominant frequency and power of each band by integrating the power spectral densities (PSD) obtained by using Welch's algorithm, and ratio of power within the low-frequency and high-frequency band (LF/HF). Since the parasympathetic activity dominates at high frequency, the ratio of LF/HF is generally thought to distinguish sympathetic effects from parasympathetic effects (Malliani, 1999).

4.2.2 Respiration (RSP)

Including the typical statistics of the raw RSP signal we calculated similar types of features like the ECG features, power mean values of three subbands (obtained by dividing the Fourier coefficients within the range 0-0.8 Hz into non-overlapped three subbands with equal bandwidth), and the set of subband spectral entropies (SSE).

In order to investigate inherent correlation between respiration rate and heart rate, we considered a novel feature content for the RSP signal. Since RSP signal exhibits quasi periodic waveform with sinusoidal property, it is not unreasonable to process HRV like analysis for the RSP signal, i.e. to estimate breathing rate variability (BRV). After detrending with mean value of the entire signal and low-pass filtering, we calculated the BRV by detecting the peaks in the signal using the maxima ranks within each zero-crossing. From the BRV time series, similar to the ECG signal, we calculated mean value, SD, SD of first difference, MSE, Poincaré analysis, etc. In the spectrum of the BRV, peak frequency, power of two subbands, low-frequency band (0-0.03Hz) and high-frequency band (0.03-0.15 Hz), and the ratio of power within the two bands (LF/HF) are calculated.

4.2.3 Skin Conductivity (SC)

The mean value, standard deviation, and mean of first and second derivations are extracted as features from the normalized SC signal and the low-passed SC signal using 0.2 Hz of cutoff frequency. To obtain a detrended SCR (skin conductance response) waveform without DC-level components, we removed continuous, piecewise linear trend in the two low-passed signals, i.e., very low-passed (VLP) with 0.08 Hz and low-passed (LP) signal with 0.2 Hz of cutoff frequency, respectively (see Fig. 6 (a)-(e)).

The baseline of the SC signal was calculated and subtracted to consider only relative amplitudes. By

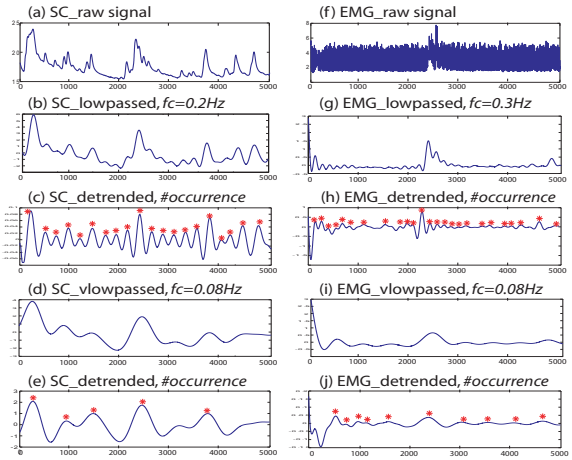


Figure 6: Analysis Examples of SC and EMG signals.

finding two consecutive zero-crossings and the maximum value between them, we calculated the number of SCR occurrences within 100 seconds from each LP and VLP signal, mean of the amplitudes of all occur-

rences, and ratio of the SCR occurrences within the low-passed signals (VLP/LP).

4.2.4 Electromyography (EMG)

For the EMG signal we calculated similar types of features as in the case of the SC signal. From normalized and low-passed signals, the mean value of entire signal, the mean of first and second derivations, and the standard deviation are extracted as features. The occurrence number of myo-responses and ratio of that within VLP and LP signals are also added in feature set by similar manner used for detecting the SCR occurrence but with 0.08 Hz (VLP) and 0.3 Hz (LP) of cutoff frequency (see Fig. 6.(f)-(j)).

Finally we obtained a total of 110 features from the 4-channel biosignals; 53 (ECG) + 37 (RSP) + 10 (SC) + 10 (EMG).

5 CLASSIFICATION RESULT

For classification we used the pseudoinverse linear discriminant analysis (pLDA) (Ye and Li, 2005), combined with the sequential backward selection (SBS) (Kittler, 1986) to select significant feature subset. The pLDA is a natural extension of classical LDA by applying eigenvalue decomposition to the scatter matrices, in order to deal with the sigularity problem of LDA. Table 1 with confusion matrix presents the correct classification ratio ($\mathcal{CCR}$) of subject-dependent (Subject A, B, and C) and subject-independent (All) classification where the features of all the subjects are simply merged and normalized. We used leave-one-out cross-validation where a single observation taken from the entire samples is used as the test data and the remaining observations are used for training the classifier. This is repeated such that each observation in the samples is used once as the test data. In the Table, it turned out that the $\mathcal{CCR}$ is depending on subject to subject. For example, best accuracy was 91% from subject B and lowest was 81% from subject A. Not only the overall accuracy but the $\mathcal{CCR}$ of the single emotions differs from subject to subject. On the other side, it is very meaningful that relatively robust recognition accuracy is achieved for the classification between emotions that are reciprocal with each other regarding the diagonal quadrants in the 2D emotion model, i.e., joy vs. sadness and anger vs. pleasure. Moreover, the accuracy is much better than that of arousal classification.

The $\mathcal{CCR}$ of subject-independent classification was not comparable to that obtained for subject-dependent classification. As shown in Figure 7, merging the features of all subjects does not refine the discriminating information related to the emotions, but rather leads to scattered class boundaries.

Table 1: Recognition results in rates (*error 0.00 = $\mathcal{CCR}$ 100%*) achieved by using pLDA with SBS and leave-one-out cross validation.
# of samples: 120 for each subject and 360 for All.

Subject A ($\mathcal{CCR}$ % = ***81%***)

	joy	anger	sadness	pleasure	*total**	*error*
joy	***22***	4	1	3	*30*	*0.27*
anger	3	***26***	1	0	*30*	*0.13*
sadness	1	2	***23***	4	*30*	*0.23*
pleasure	3	0	1	***26***	*30*	*0.13*

Subject B ($\mathcal{CCR}$ % = ***91%***)

	joy	anger	sadness	pleasure	*total**	*error*
joy	***27***	3	0	0	*30*	*0.10*
anger	3	***25***	1	1	*30*	*0.17*
sadness	0	2	***28***	0	*30*	*0.07*
pleasure	0	1	0	***29***	*30*	*0.03*

Subject C ($\mathcal{CCR}$ % = ***89%***)

	joy	anger	sadness	pleasure	*total**	*error*
joy	***28***	0	2	0	*30*	*0.07*
anger	0	***30***	0	0	*30*	*0.00*
sadness	0	0	***24***	6	*30*	*0.20*
pleasure	0	0	5	***25***	*30*	*0.17*

All: Subject-independent ($\mathcal{CCR}$ % = ***65%***)

	joy	anger	sadness	pleasure	*total**	*error*
joy	***62***	9	8	11	*90*	*0.31*
anger	15	***57***	13	5	*90*	*0.37*
sadness	9	6	***58***	17	*90*	*0.36*
pleasure	8	5	21	***56***	*90*	*0.38*

*: Actual total # of samples

We also tried to differentiate the emotions based on the two axes, arousal and valence, in the 2D emotion model. The samples of four emotions are divided into groups of negative valence (anger/sadness) and positive valence (joy/pleasure) and into groups of high arousal (joy/anger) and low arousal (sadness/pleasure). By using the same methods, we then performed two-class classification of the divided samples for arousal and valence separately. Finally, it turned out that emotion-relevant ANS specificity can be observed more conspicuously in the arousal axis regardless of subject-dependent or independent case. Classification of arousal achieved an acceptable $\mathcal{CCR}$ of 97-99% for the subject-dependent recognition and 89% for the subject-independent recognition, while the results for valence were 88-94% and 77%, respectively.

Table 2: Best emotion-relevant features extracted from four channel physiological signals. Arousal classes: joy+anger/sadness+pleasure, Valence classes: joy+pleasure/anger+sadness, Four classes: joy/anger/sadness/pleasure.

Classes	**Best Emotion-relevant Features** *(Ch_value_domain, C: ECG, R: RSP, S: SC, M: EMG)*
Arousal	*C_std(diff)_HRVtime, C_**sd2_PoincareHRV**, C_powerLow_HRVspec, R_**meanEnergy_SubSpectra**, R_**sd2_PoincareBRV** R_**mean_MSE**, S_mean_RawLowpassed, S_std_RawLowpassed, M_#occurrenceRatio_RawLowpassed M_mean_RawNormed*
Valence	*C_**sd2_PoincareHRV**, C_meanEnergy_SubSpectra, C_ratioLH_HRVspec, C_mean_MSE, C_mean(diff)_MSE R_**meanEnergy_SubSpectra**, R_mean(diff)_SubSpectra, R_sd1_PoincareBRV, R_**sd2_PoincareBRV**, R_**mean_MSE** S_mean(diff)_RawNormed, M_mean(diff)_RawNormed*
Four Emotions	*C_mean_HRVtime, C_std_HRVtime, C_std(diff)_HRVtime, C_mean(diff)_MSE, C_mean_MSE, C_mean_SSE C_**sd2_PoincareHRV**, C_mean_SubSpectra, R_**meanEnergy_SubSpectra**, R_mean_SSE, R_mean_BRVtime R_sd1_PoincareBRV, R_**sd2_PoincareBRV**, R_**mean_MSE**, R_power_BRVspec, S_std_RawLowpassed S_mean(diff)_RawNormed, S_mean(diff(diff))_RawLowpassed, S_mean_RawNormed, S_#occurrence_RawLowpassed M_mean(diff)_RawNormed*

: overall selected features are printed in bold

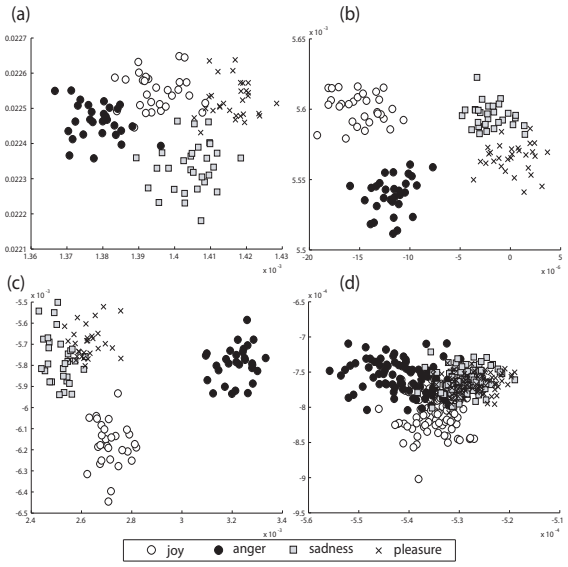


Figure 7: Comparison of feature distributions of subject-dependent and subject-independent case. (a) Subject A, (b) Subject B, (c) Subject C, (d) Subject-independent.

6 BEST EMOTION-RELEVANT ANS FEATURES

In Table 2, the best emotion-relevant features, that we determined by ranking the features selected for all subjects (including Subject All) in each classification problem, are listed in detail by specifying their values and domains. One interesting result is that each classification problem respectively links together with certain feature domain. The features obtained from time/frequency analysis of HRV time series are decisive for arousal and four emotions classification, while the features from MSE domain of ECG signals are a predominant factor for correct valence differentiation. Particularly, mutually sympathizing correlate between HRV and BRV which is firstly proposed in this paper has been clearly observed in all the classification problems by the features from their time/frequency analysis and Poincaré domain, *_PoincareHRV* and *_PoincareBRV*. This reveals a manifest cross-correlation between respiration and cardiac activity with respect to emotional state. Furthermore, the correlation between heart rate and respiration is obviously captured by the features from HRV power spectrum (*_HRVspec*), the fast/long-term HRV/BRV analysis using Poincaré method, and the multiscale variance analysis of HRV/BRV (*_MSE*), and that the peaks of high frequency range in HR subband spectrum (*_SubSpectra*) provide information about how the sinoatrial node responds to vagal activity at certain respiration frequency.

In addtion, we analyzed the number of selected features for the three classification problems, arousal, valence, and four emotion states. For the arousal classification, relatively few features are used but achieved higher recognition accuracy compared to the other class problems. After the ratio of number of selected features to the total feature number of each channel, it was obvious that the SC and EMG activities reflected in both *_RawLowpassed* and *_RawNormed* domains (see Table 2) are more significant for arousal classification than the other channels. This supports also the experimental elucidation in previous works that the SCR is linearly correlated with the intensity of arousal. On the other side, we could observe a remarkable increase of number of the ECG and RSP features for the case of valence classification.

7 CONCLUSIONS

In this paper, we treated all essential stages of automatic emotion recognition system using multichannel physiological measures, from data collection up to classification process, and analyzed the results from each stage of the system. For four emotional states of three subjects, we achieved average recognition accuracy of 91% which connotes more than a prima facie evidence that there exist some ANS differences among emotions.

A wide range of physiological features from various analysis domains including time, frequency, entropy, geometric analysis, subband spectra, multiscale entropy, and HRV/BRV has been proposed to search the best emotion-relevant features and to correlate them with emotional states. The selected best features are specified in detail and their effectiveness is proven by classification results. We found that SC and EMG are linearly correlated with arousal change in emotional ANS activities, and that the features in ECG and RSP are dominant for valence differentiation. Particularly, the HRV/BRV analysis revealed the cross-correlation between heart rate and respiration.

As we humans use several modalities jointly to interpret emotional states since emotion affects almost all modes, one most challenging issue in near future work is to explore multimodal analysis for emotion recognition. Toward the human-like analysis and finer resolution of recognizable emotion classes, an essential step would be therefore to find innate priority among the modalities to be preferred for each emotional state. In this sense, physiological channel can be considered as a "baseline channel" in designing a multimodal fashion of emotion recognition system, since it provides several advantages over other external channels and acceptable recognition accuracy, as we presented in this work.

ACKNOWLEDGEMENTS

This research was partially supported by the European Commission (HUMAINE NoE: FP6 IST-507422).

REFERENCES

Costa, M., Goldberger, A. L., and Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev.*, E 71(021906).

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Mag.*, 18:32–80.

Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9:87–108.

Healey, J. and Picard, R. W. (1998). Digital processing of affective signals. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 3749–3752, Seattle, WA.

Kamen, P. W., Krum, H., and Tonkin, A. M. (1996). Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity. *Clin. Sci.*, 91:201–208.

Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42:419–427.

Kittler, J. (1986). *Feature Selection and Extraction*, pages 59–83. Academic Press, Inc.

LeDoux, J. E. (1992). *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, pages 339–351. New York: Wiley-Liss.

Malliani, A. (1999). The pattern of sympathovagal balance explored in the frequency domain. *News Physiol. Sci.*, 14:111–117.

Nasoz, F., Alvarez, K., Lisetti, C., and Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work - Special Issue on Presence*, 6(1).

Pan, J. and Tompkins, W. (1985). A real-time qrs detection algorithm. *IEEE Trans. Biomed. Eng.*, 32(3):230–323.

Picard, R., Vyzas, E., and Healy, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. and Machine Intell.*, 23(10):1175–1191.

Richmann, J. and Moorman, J. (2000). Physiological time series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol. 278*, H2039.

Ye, J. and Li, Q. (2005). A two-stage linear discriminant analysis via qr-decomposition. *pami*, 27(6).

AN ECoG BASED BRAIN COMPUTER INTERFACE WITH SPATIALLY ADAPTED TIME-FREQUENCY PATTERNS

Nuri F. Ince, Fikri Goksu and Ahmed H. Tewfik
Department of Electrical and Computer Engineering, University of Minnesota, 200 Union St. SE, 55455, Minneapolis, U.S.A.
firat@umn.edu, goks0002@umn.edu, tewfik@umn.edu

Keywords: Electrocorticogram, Brain Computer Interface, Time Frequency, Undecimated Wavelet Packet Transform.

Abstract: In this paper we describe an adaptive approach for the classification of multichannel electrocorticogram (ECoG) recordings for a Brain Computer Interface. In particular the proposed approach implements a time-frequency plane feature extraction strategy from multichannel ECoG signals by using a dual-tree undecimated wavelet packet transform. The dual-tree undecimated wavelet packet transform generates a redundant feature dictionary with different time-frequency resolutions. Rather than evaluating the individual discrimination performance of each electrode or candidate feature, the proposed approach implements a wrapper strategy to select a subset of features from the redundant structured dictionary by evaluating the classification performance of their combination. This enables the algorithm to optimally select the most informative features coming from different cortical areas and/or time frequency locations. We show experimental classification results on the ECoG data set of BCI competition 2005. The proposed approach achieved a classification accuracy of 93% by using only three features.

1 INTRODUCTION

Brain-computer interfaces (BCIs) use the electrical activity of the brain for communication and control. Since the muscles are bypassed, a BCI can be used by people with motor disabilities to interact with their environment. Electroencephalogram (EEG) is widely used in BCIs due to its non-invasiveness. However, the low signal to noise ratio (SNR) and spatial resolution of EEG limit its effectiveness in BCIs. On the other hand invasive methods such as single neuron recordings have higher spatial resolution and SNR. However, they have clinical risks. Furthermore, maintaining long term reliable recording with implantable electrodes is difficult. On the other hand, an electrocorticogram (ECoG) has the ability to provide long term recordings from the surface of brain. Furthermore, ECoG signals also provide information about oscillatory activities in the brain with a much higher bandwidth than EEG (Leuthardt 2004). Therefore, existing algorithms for EEG classification are readily applicable to ECoG processing.

Various events in brain signals such as slow cortical potentials, motor imagery (MI) related sensorimotor rhythms, and visual evoked potentials were used in construction of ECoG based BCIs (Wolpaw 2000, Pfurtscheller 2001). In MI based BCIs, the subjects are asked to perform an imagined rehearsal of either hand/finger or foot movement without any muscular output. Related events in sensorimotor rhythms such as alpha (7-13Hz) and beta (16-32Hz) bands are processed to recognize the executed task using only brain waves. Several methods have been proposed to extract relevant features for BCI classification from rhythmic activities. Methods such as autoregressive modeling and sub band energies in predefined windows are widely used in single trial ECoG classification (Schlogl 1997, Prezenger 1999). When used with multi channel recordings, all of these methods need to deal with the high dimensionality of the data. Selecting the most informative electrodes and adapting to subject specific oscillatory patterns is critical for accurate classification. However, due to the lack of prior knowledge, selection of the most informative electrode locations can be difficult. Furthermore, it is well known that there exists a great deal of inter subject variability of EEG and ECoG patterns in spatial, temporal, and frequency domains (Ince 2006, Ince 2007, Leuthardt 2004, Prutscheller 2001 and Schlogl 1999). In (Ramoser 2000), the common spatial patterns (CSP) method was proposed to classify multichannel EEG recordings. The CSP

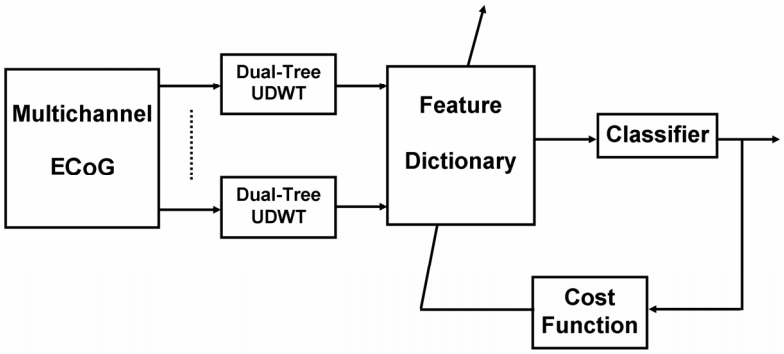


Figure 1: The block diagram of the proposed feature extraction and feature subset selection technique.

method weighs each electrode location for classification and uses the correlation between channels to increase the SNR of the extracted features. Although the performance is increased with CSP, it has been shown that this method requires a number of electrodes to improve classification accuracy and that it is very sensitive to electrode montage. Furthermore, since it uses the variance of each channel, this method does not account for the spatiotemporal differences in distinct frequency subbands. Recently, time-frequency methods have been proposed as an alternative strategy for the extraction of MI related patterns in BCI's (Wang 2004, Ince2006 and Ince 2007). These methods utilized the entire time-frequency plane of each channel and integrate components with different temporal and spectral characteristics. Promising results were reported on well known data sets while classifying multichannel EEG. One of the main difficulties with these methods is once again dealing with the high dimensionality of the data collected. Furthermore, the adaptation to important patterns is implemented either by only accounting for the discrimination power of individual electrode locations or simultaneous processing of a large number of electrodes.

In this paper we tackle these problems by implementing a spatially adapted time-frequency plane feature extraction and classification strategy. To our knowledge this is the first time that an approach implements a joint processing of ECoG features with different time and frequency resolution coming from distinct cortical areas for classification purposes. The algorithm proposed in this paper requires no prior knowledge of relevant time-frequency indices and related cortical areas. In particular, as a first step, the proposed approach implements a time-frequency plane feature extraction strategy on each channel from multichannel ECoG signals by using a dual-tree undecimated wavelet packet transform (UDWT). The dual-tree undecimated wavelet packet transform forms a redundant, structured feature dictionary with different t-f resolutions. In the next step, this redundant dictionary is used for classification. Rather than evaluating the individual discrimination performance of each electrode or candidate feature, the proposed approach selects a subset of features from the redundant structured dictionary by evaluating the classification performance of their combination using a wrapper strategy. This enables the algorithm to optimally select the most discriminative features coming from different cortical areas and/or time-frequency locations. A block diagram summarizing the technical concept is given in Figure 1. In order to evaluate the efficiency of the proposed method we test it on the ECoG dataset of BCI competition 2005.

The paper is organized as follows. In the next section we describe the extraction of structural time-frequency features with dual-tree undecimated wavelet transform. In the following section we discuss available feature selection procedures and details of our proposed solutions. We describe the multichannel ECoG data in section 4. Finally we provide experimental results in section 5 and discuss our findings in section 6.

2 FEATURE EXTRACTION

Let us describe our feature dictionary and explain how it is computed from the wavelet-based dual-tree structure. A schematic diagram of the dual tree is shown in Figure 2. As indicated in the previous sections, the ECoG can be divided into several

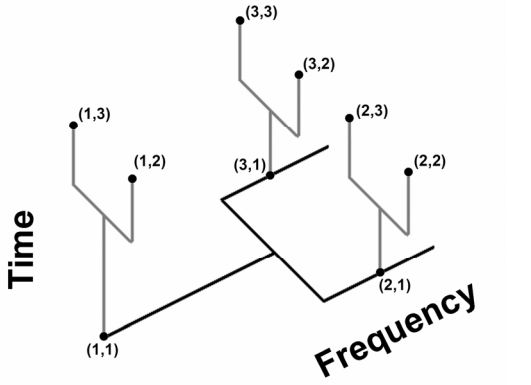


Figure 2: This dual tree uses 1-level in both planes. Each node of the horizontal tree is a frequency subbands. Node {1,1} represents unfiltered original signal, node{2,1} represents low pass filtered signal and node {3,1} high pass filtered. Each of these subbands is segmented in time into 3 segments, as shown in the vertical tree. Segment {1,1},{2,1} and {3,1} covers the whole subband, segment. {1,2},{2,2} and {3,2} covers the first and segments with time indices three the second half of it.

frequency subbands with distinct and subject depended characteristic. In order to extract information from these rhythms, we examine subbands of the ECoG signal by using an undecimated wavelet transform. In each subband, a second pyramidal tree is utilized to extract the time varying characteristics of the subband.

2.1 Undecimated Wavelet Transform

Discrete Wavelet Transform (DWT) and its variants have been extensively used in 1D and 2D signal analysis (Vetterli 2001). However, the downsampling operator at the outputs of each filter produces a shift variant decomposition. In practice, a shift in the signal is reflected by abrupt changes in the extracted expansion coefficients or related features. In (Unser 1995) the undecimated wavelet transform is proposed to extract subband energy features which are shift invariant. This is achieved by removing the downsampling operation. The output at any level of pyramidal filter bank is computed by using an appropriate filter which is derived by upsampling the basic filter.

A filter *g(n)* with a z-transform *G(z)* that satisfies the quadrature mirror filter condition

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1 \qquad (1)$$

is used to construct the pyramidal filter bank (Figure 3). The high-pass filter *h(n)* is obtained by shifting

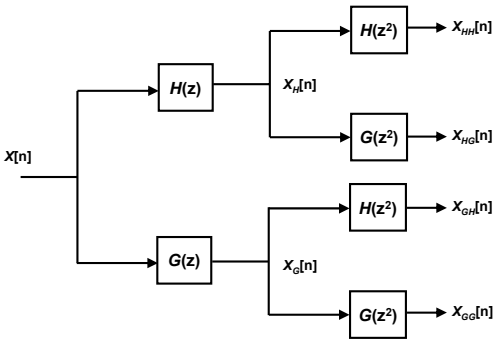


Figure 3: The pyramidal undecimated wavelet tree.

and modulating *g(n)*. Specifically, the z transform of *h(n)* is chosen as

$$H(z) = zG(-z^{-1}). \qquad (2)$$

The subsequent filters in the filter bank are then generated by increasing the width of *f(n)* and *g(n)* at every step, e.g.,

$$G_{i+1}(z) = G(z^{2^i})$$
$$H_{i+1}(z) = H(z^{2^i}), \text{ (i=0,1,. . . ., N).} \qquad (3)$$

In the signal domain, the filter generation can be expressed as

$$g_{i+1}(k) = [g]_{\uparrow 2^i}$$
$$h_{i+1}(k) = [h]_{\uparrow 2^i} \qquad (4)$$

where the notation $[]_{\uparrow m}$ denotes the up-sampling operation by a factor of m.

The horizontal pyramidal tree of Fig.2 provides subband decomposition of the ECoG signal. Next, we segment the signal in each subband with rectangular time windows. Such an approach will extract the temporal information in each subband. As in the frequency decomposition tree, every node of the frequency tree is segmented into time segments with a pyramidal tree structure. Each parent time window covers a space as the union of its children windows. In a given level, the length of a window is equal to $L/2^t$ where L is the length of signal and t denotes the level. The time segmentation explained above forms the second branch (vertical) of the double tree. After segmenting the signal in time and frequency, we retain the energy of each node of the dual-tree as a feature. By using a dual tree structure we could calculate a rich library of features describing the ECoG activities with several

spectro-temporal resolutions. From now on we keep the index information of the dual tree structure to be used in the later stage for dimension reduction via pruning.

To summarize this section the reader is referred to the double tree structure in Fig. 2. Note that the dual tree structure satisfies two conditions:

- For a given node in the frequency tree, the mother band covers the same frequency band width (BW) as the union of its children

$$BW_{Mother} \supset (BW_{Child1} \cup BW_{Child2}) \quad (5)$$

- This same condition is also satisfied along the time axis. For a given node, the number of time samples (TS) of the mother window is equal to that of the union of its children.

$$TS_{Mother} \supset (TS_{Child1} \cup TS_{Child2}) \quad (6)$$

These two properties allow us to prune the tree structure. When a particular feature index is selected, one can remove those indices from the dual tree structure that overlap in time and frequency with the selected index. Let $\boldsymbol{T}$ be the number of levels use to decompose the signal in time and $\boldsymbol{F}$ be the number levels use to decompose the signal in the frequency domain, there will be $\mathbf{2^{(F+1)}-1}$ subbands (including the original signal) and $\mathbf{2^{(T+1)}-1}$ time segments for each subband. This will make the total number of potential features $\mathbf{NF=(2^{(F+1)}-1)(2^{(T+1)}-1).}$

3 SUBSET SELECTION

Calculating the dual-tree features for each electrode location forms a redundant feature dictionary. The redundancy comes from the dual tree structure. As explained in the previous section the dual tree has total $\boldsymbol{NF}=\mathbf{(2^{(F+1)}-1)(2^{(T+1)}-1)}$ features for each signal where $\boldsymbol{F}$ is the total number of frequency levels and $\boldsymbol{T}$ the total number of time levels. In a typical case, T=3, F=4 and over 64 electrodes are used resulting in a dictionary with around thirty thousand features. In such a high dimensional space (NF=29760) the classifier may easily go into over-learning and provide a lower generalization capability.

Here, we incorporate the structural relationship between features in the dictionary and use several feature subset selection strategies to reduce the dimensionality of the feature set. Since the features are calculated in a tree structure, efficient algorithms were proposed in the past for dimensionality reduction. In (Saito 1996) a pruning approach was proposed which utilizes the relationship between the mother and children subspaces to decrease the dimensionality of the feature set. In particular, each tree is individually pruned from bottom to top by maximizing a distance function. The resulting features are sorted according to their discrimination power and the top subset is used for classification. Although such a filtering strategy with pruning will provide significant dimension reduction by keeping the most predictive features, it does not account for the interrelations between features in the final classification stage. Here, we reshape and combine the pruning procedure for feature selection with a wrapper strategy. In particular, we quantify the efficiency of each feature subset by evaluating its classification accuracy with a cost measure and we use this cost to reformulate our dictionary via pruning.

Four different types of methods are considered for feature selection in this study. The structure in Figure 1 is general representation of each of the four methods. The left most box in Figure 1 is the rich time-frequency feature dictionary. On the right end a linear discriminant (LDA) is used both for classification and extracting the relationship among combinations of features. This output is fed to a cost function to measure the discrimination power for that combination of features. This measure will be used to select the best among all other feature combinations. Furthermore, depending on the selected feature index, a pruning operation will be implemented to reduce the dimensionality in the rich feature dictionary.

In this particular study, the Fisher Discrimination (FD) criterion is used as a cost function.

$$FD = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}. \quad (7)$$

The four different strategies mentioned above are: Sequential forward feature selection (SFFS), SFFS with pruning (SFFS-P), Cost function based pruning and feature Selection (CFS), and CFS with principal component analysis (PCA) post processing.

3.1 Sequential Forward Feature Selection: SFFS

The SFFS is a wrapper strategy which selects a subset of features one by one. A cost function is used on classifier output to measure the efficiency of each feature. By using LDA, the feature vectors are projected on a one dimensional space. Then the FD criterion was used to estimate the efficiency of the projection. After this search is done over all feature vectors, the best feature index is selected by

comparing the cost values of each feature vector. In the next step the feature vector which will do the best in combination with the first selected ones is identified by searching over the remaining feature vectors. This procedure is run until a desired number of features is reached. Note that SFFS uses all the boxes and connections in Figure 1 except for the feedback from the cost function to the dictionary. Since no dimension reduction is implemented on the entire feature space, this approach has high computational complexity.

3.2 SFFS with Pruning: SFFS-P

The SFFS-P is also a wrapper strategy with an additional pruning module for dimension reduction. Once a feature index is selected, the corresponding frequency tree and time tree indexes are calculated on the dual-tree. Then the nodes that overlap with the selected feature index in time and frequency are removed. Next, the feature which will do best in combination with the first selected feature is identified by searching the pruned dictionary. In other words, the dictionary is pruned based on the last selected feature. This procedure is run until the desired number of features is reached. Therefore, the only difference between SFFS and SFFS-P is that pruning is done on the dictionary based on the selected features. This provides a fast decrease in the number of candidate features and complexity is much smaller than SFFS.

3.3 Cost Function based Pruning and Feature Selection (CFS)

The CFS is a filtering approach that uses the structure in the feature dictionary for pruning. After finalizing the pruning procedure for each electrode location, it uses a cost function to rank the features. In particular, it uses the FD criterion to rank the features. It does not use either the LDA or the feedback path in Figure 1. Instead, using the FD measure, a cost value is computed for each node on the double tree individually. Then a pruning algorithm is run on the double tree by keeping the nodes with maximum discrimination. Once a node is selected all nodes overlapping with the selected one are removed. This procedure is iterated until no pruning can be implemented. After pruning the dual-trees for each electrode location, the resulting feature set is sorted according to their corresponding discrimination power and input to the classifier. In this way the most predictive features were entered to the classification module. Since no feedback is used from the classifier, the CFS has lower computational complexity than the other two methods.

The CFS method works as a filter on the electrodes by only keeping those indices with maximum discrimination power. However, since features are evaluated according to their discrimination power individually, such a method does not account for the correlations between features. In (Ince 2006 and Ince 2007) PCA analysis is performed on a subset of top sorted features to obtain a decorrelated feature set. The PCA post processed features are sorted according to their corresponding eigenvalues in decreasing order and used in classification. Here we will also use the PCA as a post processing step with the CFS to obtain a deccorelated feature set. We will refer this method as CFS-PCA.

4 MULTICHANNEL ECoG DATA

In order to evaluate the performance of the proposed method we used the multichannel ECoG (Lal 2005) dataset of BCI competition 2005 (ida.first.fraunhofer.de/projects/bci/competition_iii/) During the BCI experiment, a subject had to perform imagined movements of either the left small finger or the tongue. The ECoG data was recorded using an 8x8 ECoG platinum electrode grid which was placed on the contralateral (right) motor cortex as shown in Figure 4. All recordings were performed with a sampling rate of 1000Hz. Every trial consisted of either an imagined tongue or an imagined finger movement and was recorded for 3 seconds duration. To avoid visually evoked potentials being reflected

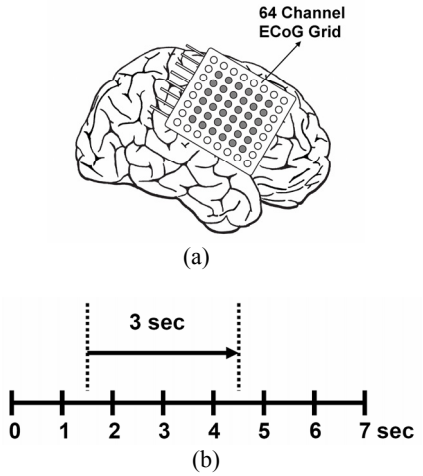


Figure 4: The 8x8 electrode grid was placed on the right hemisphere over the motor cortex (Modified from Lal 2005). For surface Laplacian derivation only marked electrodes are used. (b) The timing diagram of the experimental paradigm. The go cue for motor imagery is given at second one. A three second time window starting after 500ms of go cue is used to classify ECoG data.

by the data, the recording intervals started 0.5 seconds after the visual cue had ended. Each channel was filtered with a low pass filter in 0-120Hz band. The filtered data was down sampled by a factor 4 to 250Hz. Each trial was expanded from 750 samples into 768 samples by symmetric extension on the right side to enable segmentation in a pyramidal tree structure. Besides monopolar data, we also consider ECoG data that is processed using a surface Laplacian derivation. More specifically, each electrode data is subtracted from the weighted average of the surrounding 6 electrodes. The electrodes on the border are eliminated from the analysis resulting in a total of 36 electrodes (See Figure 4). For monopolar data all 64 electrodes were used for analysis. We used 278 trials for training and 100 trials for testing. The training and test data were recorded from the same subject and with the same task, but on two different days with about 1 week in between.

5 RESULTS

To extract the dual tree features we select T=3 and F=4. For a 125 Hz bandwidth, the frequency tree provided around 8Hz resolution at the finest level. Along the time axis, the time resolution was 375ms. The 12 tap Daubechies filter (db6) was used in constructing the frequency tree of the UDWT. In order to learn the most discriminant time-frequency indices and the corresponding cortical areas we utilized a 10 times 10 fold cross validation in the training dataset. The optimal feature number at which the classification error is minimal is selected from the averaged cross validation error curves. Then, the learned feature indices are used in testing the classifier on the test set. The results obtained with the different methods are presented in Table.1.

We note that the SFFS and SFFS-P provided the highest classification accuracy with only three features on the test set using the Laplacian derivation. Although a lower error rate was achieved by CFS with the training data, interestingly, the testing error rate of the CFS was higher than those of the other methods. We also note that a large number of features were used by CFS to achieve 9.9% error rate in the training set. In contrast, the SFFS and SFFS-P algorithms used only 3 features to achieve the minimum 10.2% error rate. The cross validation error curves versus the number of features are given in Figure 5. Since the results using Laplacian derivation outperformed those obtained with monopolar data, only the results corresponding to the former are provided.

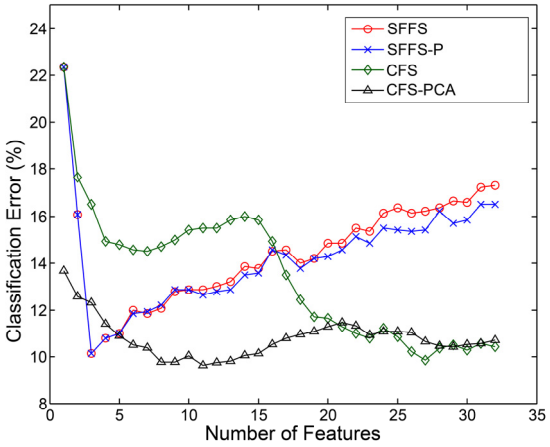


Figure 5: The cross validation error curves for the different methods in the training data.

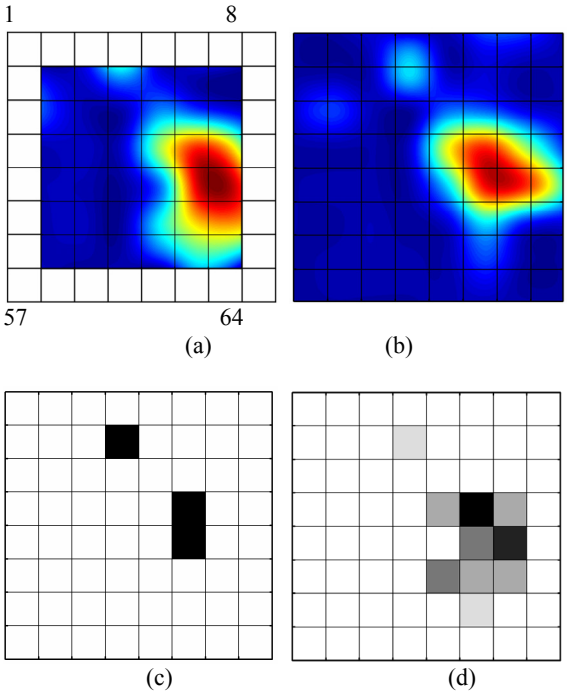


Figure 6: Discriminant cortical areas (a) Laplacian (b) Monopolar. The number of selected features from different electrode locations in Laplacian derivation for SFFS-P (c) and for CSF(d) are given. The darker areas indicate a higher number of features are selected from these regions. Note that SFFS-P provides a balanced feature distribution. The CSF selected most of 27 features from the same region.

As can be seen clearly from these curves, SFFS and SFFS-P select the best combination of features and achieve the minimal error with only three features.

Table 1: The cross validation (CV) and test error rates of different methods and related number of features (NoF) used for final classification.

		Training		Test
	Method	***CV Error (%)***	***NoF***	***Error (%)***
Laplacian	**SFFS**	**10.2**	**3**	**7**
	SFFS-P	**10.2**	**3**	**7**
	CSF	9.9	27	18
	CSF-PCA	9.6	11	8
Monopolar	SFFS	12.6	3	20
	SFFS-P	10.3	4	9
	CSF	11.7	22	12
	CSF-PCA	11.2	14	8

Furthermore, using the structure of the feature dictionary, SFFS-P achieves this result with reduced complexity due to pruning. The pruning process provides a dimension reduction and feature decorrelation. CFS, on the other hand, achieves the minimal error using a large number of features. The interactions among the selected features cannot be taken into account with this approach. In addition the correlated neighbor areas may result in a duplication of information in the sorted features. In order to decorrelate the features a Principal Component Analysis (PCA) was employed on the CFS ordered features. This post processing step provided lower error rates than those achieved by CFS alone. The test error rate was 8% for the PCA post processed features. It should be noted here that CFS-PCA produced comparable results with those of SFFS and SFFS-P. However one should note that PCA induces an additional complexity. This method requires all 32 features to be extracted from ECoG which leads to a much higher computational complexity compared to three features selected by SFFS and SFFS-P.

Since the testing data was recorded on another date, the variability in the ECoG signal is expected. The results obtained indicate that the CFS algorithm is very sensitive to this type of variability. Although the cross validation error in the training set was low, the testing error rate was much higher compared to other methods. We believe that the correlated activity across cortical areas is an important reason why CFS selects the same information repeatedly. Since the SFFS and SFFS-P have the advantage of examining the interactions between different cortical areas and t-f locations, these subset selection algorithms can form a more effective subset of features for classification. In order to support our hypothesis we show the discriminatory cortical maps of monopolar and Laplacian derivations in Figure 6. In order to generate these images we used the most discriminant feature of each electrode location and produced an image over the 8x8 grid to present the distribution of the most discriminative locations. Furthermore, we mark the electrode locations selected by SFFS, SFFS-P, and CFS for classification. After inspecting Figure 6 (a) and (b) we noticed that a large number of neighbor electrode locations carry discriminant information. The CFS method used a large number of electrodes from this region for classification. In contrast, the SFFS and SFFS-P methods selected another cortical area from upper side of the grid. Even though this electrode location does not seem to be very discriminative, it played a key role in achieving a lower classification rate on the validation data.

Since only three features are used by SFFS and SFFS-P, they are more robust to intra-subject variability of ECoG signals. Note also that the error rate in monopolar derivation is much higher than that of the Laplacian derivation. We observed large DC changes in ECoG signals in the test data set. Since the Laplacian derivation provides a differential operator, large baseline wanders affecting many electrodes are eliminated in this setup. However, for the monopolar recordings the features are very sensitive to this type of changes.

Note also that the validation accuracy of SFFS and SFFS-P in the test set is higher than the cross validation accuracy. One of the underlying reasons could be that the subject can control his/her brain patterns with a higher accuracy with the increasing number of trials. In addition the SNR of the signals might have improved over time due to tissue electrode interaction.

Finally, we compared our proposed method's test result with those of achieved at the BCI competition in 2005 using the same ECoG data. The classification accuracies and methods used in each method are presented in Table 2. Our method achieved the best result of 7% error with both SFFS

Table 2: The comparison of the proposed method with the best three methods from the BCI 2005 competition.

Features Used	Classifier	Error (%)
UDWT based subband energies	LDA	7
Common Spatial Subspace Decomposition	Linear SVM	9
ICA combined with spectral power and AR coefficients	Regularized logistic regression	13
Spectral power of manually selected channels	Logistic regression	14

and SFFS-P methods. We note that our proposed approach has outperformed both CSP and AR model based techniques.

6 CONCLUSIONS

In this paper we proposed a new feature extraction and classification strategy for multi-channel ECoG recordings in a BCI task. Rather than using predefined frequency indices or manually selecting cortical areas, the algorithm implemented an automatic feature extraction and subset selection procedure over a redundant time-frequency feature dictionary. This feature dictionary was obtained by decomposing the ECoG signals into subbands with an undecimated wavelet transform and then segmenting each subbband in time successively. By combining a wrapper strategy with dictionary pruning, the method achieved 93% classification accuracy using only three features. The results we obtained show that the proposed method is a good candidate for the construction of an ECoG based invasive BCI system with very low computational complexity and high classification accuracy.

REFERENCES

Ince N. F., Tewfik A., Arica S., 2007, "Extraction subject-specific motor imagery time-frequency patterns for single trial EEG classification", *Comp. Biol. Med. Elsevier.*

Ince N. F., Arica S., Tewfik A., 2007, "Classification of single trial motor imagery EEG recordings by using subject adapted non-dyadic arbitrary time-frequency tilings," *J. Neural Eng. 3, 235-244.*

Lal Thomas N., et.al., *2005,* Methods Towards Invasive Human Brain Computer Interfaces*. Advances in Neural Information Processing Systems (NIPS17, 737-744. (Eds.) Saul, L. K., Y. Weiss, L. Bottou, MIT Press, Cambridge, MA, USA).*

Leuthardt E. C. , Schalk G., Wolpaw J. R., Ojemann J. G. and Moran D. W., 2004, A brain–computer interface using electrocorticographic signals in humans, *Journal of Neural Engineering, pp. 63–71.*

Pfurtscheller G., Neuper C., 2001. Motor Imagery and Direct Brain-Computer Interface. *Proceedings of IEEE, vol.89, pp. 1123-1134.*

Prezenger M., Pfurtscheller G., 1999. Frequency component selection for an EEG-based brain computer interface. *IEEE Trans. on Rehabil. Eng. 7, pp. 413-419.*

Ramoser H., Müller-Gerking J., and Pfurtscheller G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehab. Eng., vol. 8, no. 4, pp. 441–446.*

Saito N. et al., *2002,* Discriminant feature extraction using empirical probability density and a local basis library, *Pattern Recognition, vol.35, pp. 1842-1852.*

Schlögl A., Flotzinger D., Pfurtscheller G., 1997. Adaptive autoregressive modeling used for single trial EEG classification. *Biomed. Technik,42, pp. 162-167.*

Unser M., 1995. Texture classification and segmentation using wavelet frames*. IEEE Trans. Image Proc., pp. 1549–60, Vol.4(11), Nov.*

Vetterli M., 2001. Wavelets, approximation, and compression," *IEEE Signal Proc. Magazine, pp. 59–73, Sept..*

Wang T., and B. He, 2004. Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns. *Clin. Neuro. vol.115, pp. 2744–2753.*

Wolpaw J. R., et.al., 2000. Brain-Computer Interface Technology: A review of the first international meeting, *IEEE Trans. On Rehab. Eng. 8 164-73.*

A SUPERVISED LEARNING APPROACH BASED ON THE CONTINUOUS WAVELET TRANSFORM FOR R SPIKE DETECTION IN ECG

G. de Lannoy[1,2], A. de Decker[1] and M. Verleysen[1]
[1]*Machine Learning Group, Université catholique de Louvain*
pl. du Levant 3, 1348 Louvain-la-Neuve, Belgium
[2]*Departement of physiology and pharmacology, Université catholique de Louvain*
av. Hippocrate 54, 1200 Bruxelles, Belgium
gael.delannoy@uclouvain.be, arnaud.dedecker@uclouvain.be, michel.verleysen@uclouvain.be

Keywords: Continuous wavelet transform, automatic ECG annotation, R spike detection, supervised learning.

Abstract: One of the most important tasks in automatic annotation of the ECG is the detection of the R spike. The wavelet transform is a widely used tool for R spike detection. The time-frequency decomposition is indeed a powerful tool to analyze non-stationary signals. Still, current methods use consecutive wavelet scales in an a priori restricted range and may therefore lack adaptivity. This paper introduces a supervised learning algorithm which learns the optimal scales for each dataset using the annotations provided by physicians on a small training set. For each record, this method allows a specific set of non consecutive scales to be selected, based on the record characteristics. The selected scales are then used on the original long-term ECG signal recording and a hard thresholding rule is applied on the derivative of the wavelet coefficients to label the R spikes. This algorithm has been tested on the MIT-BIH arrhythmia database and obtains an average sensitivity rate of 99.7% and average positive predictivity rate of 99.7%.

1 INTRODUCTION

In the framework of biomedical engineering, the analysis of the electrocardiogram (ECG) is one of the most widely studied topics. The easy recording and visual interpretation of the non-invasive electrocardiogram signal is a powerful way for medical professionals to extract important information about the clinical condition of their patients.

The ECG is a measure of the electrical activity associated with the heart. It is characterized by a time-variant cyclic occurrence of patterns with different frequency content (QRS complexes, P and T waves). The P wave corresponds to the contraction of the atria, the QRS complex to the contraction of the ventricles and the T wave to their repolarization. Because the ventricles contain more muscle mass than the atria, the QRS complex is more intensive than the P wave. The QRS wave is therefore the most representative feature of the ECG. Furthermore, once the QRS complex has been identified, other features of interest can be more easily detected.

Analyzing ECGs for a long time can lead to errors and misinterpretations. This is the reason why automatic feature extraction of the ECG signal can help physicians in their diagnosis for early detection of cardiac troubles. The feature extraction mainly consists in the automatic annotation of the different waves in the recording, the most important of them being the QRS. One of the main application of the QRS detection is the heart rate variability (HRV) analysis (Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology, 1996). HRV measures have been proven successful in diagnosing cardiac abnormalities and neuropathies or evaluating the actions of the autonomic nervous system on the heart (Acharya et al., 2006). However, HRV measures heavily rely on the accuracy of the QRS feature detection on the digitalized ECG signal.

Automatic feature extraction and especially R spike detection is thus a milestone for ECG analysis. However, it is a difficult task in real situations: (1) The physiological variations due to the patient and its disease make the ECG a non-stationary signal. (2) Other ECG components such as the P or T wave looking like QRS complexes often lead to wrong detections. (3) There are many sources of noise that pol-

lute the ECG signal such as power line interferences, muscular artifacts, poor electrode contacts and baseline wanderings due to respiration. These three problems highly compromise the detection of R spikes.

The detection of QRS complexes in the ECG has been conducted by many researchers in the past years. However, none of the current algorithms are able to automatically learn their parameters using pre-labeled beats provided by physicians. The aim of this paper is to introduce a new algorithm for R peak detection that does not blindly detect beats but learns and propagates the annotations provided by physicians on a small portion of the signal, which is often wanted in real situations. Our contribution consists in the design and experiment of a supervised learning algorithm for an optimal and automatic signal decomposition for further optimal R spike detection. The associated detection method by hard thresholding rule is also presented. The algorithm does not require any pre-processing of the signal and can also be adapted for the detection of other features such as the P or T wave.

The following of this paper is structured as follows. After this introduction, section 2 gives a brief literature review about the state of the art on ECG feature detection and especially the QRS detection. Section 3 provides a summary of the theory about the continuous wavelet transform used in this paper. Section 4 introduces the methodology followed by the algorithm and section 5 shows the experiments and results obtained on a real public database.

2 STATE OF THE ART

Due to the non-stationarity of the ECG signal, the physiological conditions and the presence of many artifacts, finding a robust and general algorithm for ECG feature detection is a tough task. A lot of work has been published in the literature about the detection of various interesting ECG features such as P waves, QRS waves, T waves, QT intervals or abnormal beats by numerous techniques (Addison, 2005; Sahambi et al., 2000; Senhadji et al., 1995). This paper focuses on R spike detection only.

For this purpose, several approaches using different signal processing methods have been reported previously: template matching (Dobbs et al., 1984), mathematical models (Pahlm and Sornmo, 1984), signal envelop (Nygards and Sornmo, 1983), matched filters (Koeleman et al., 1985), ECG slope criterion (Algra and Zeelenberg, 1987), dynamic time warping (Vullings et al., 1998), syntactic methods (Kohler et al., 2002), hidden Markov models (Clavier et al., 2002), beat detection by neural networks (Xue et al., 1992; Shyuand et al., 2004), adaptative thresholding (Madeiro et al., 2007; Christov, 2004), time-frequency decompositions by wavelet transforms (Addison, 2005), and geometrical approach (Surez et al., 2007).

Among all these methods, the time-frequency decompositions by wavelet transform (WT) seem the most intuitive tool for ECG analysis. The WT is naturally appropriate for analyzing non-stationary signals because it allows precise time-frequency representation of the signal with a low computational complexity. A lot of work has been published in past years on the use of the WT for QRS detection. In 1995, (Li et al., 1995) used an algorithm based on finding the maxima larger than a threshold obtained from the pre-processed initial beats. Later, (Kadambe et al., 1999) produced a method allocating a R peak at a point being the local maxima of several consecutive dyadic wavelet scales. In both these methods, a post-processing allowed to eliminate false R detections. Based on these two publications, a lot of other researches were published on the beat detection based on the WT (Shyuand et al., 2004; Fard et al., 2007; Martinez et al., 2004; Addison, 2005; Chen et al., 2005; Chen et al., 2006).

The main problem of the WT is that one has to choose the mother wavelet and the scales used to analyze the signal on an empirical basis. While the mother wavelet can easily be chosen based on its characteristics and ressemblance with a QRS wave, the ideal scale(s) at which the QRS are matched is harder to guess *a priori*. Current algorithms blindly search for QRS complexes in a limited number of consecutive scales selected in a range of *a priori* fixed scales. However, the shape of the QRS pattern can be varying between patients but also with time. One or several consecutive fixed wavelet scales may not be enough to match all complexes at once in a dataset. In this paper, we propose a new supervised learning algorithm based on the continuous wavelet transform that overcomes these issues. It only relies on the annotations provided by physicians on a small portion of the signal in order to select the optimal subset of non-consecutive scales for each dataset.

3 THEORY OF THE CONTINUOUS WAVELET TRANSFORM

The continous wavelet transform (CWT) is a tool which produces a time-frequency decomposition of a

signal $x(t)$ by the convolution of this signal with a so-called *wavelet function*.

A wavelet function $\psi(t)$ is a function with several properties. It must be a function of finite energy, that is

$$E = \int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty, \quad (1)$$

and it must have a zero mean.

From a wavelet function, one can obtain a family of time-scale waveforms by translation and scaling

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\Big(\frac{t-b}{a}\Big) \quad (2)$$

where $a > 0$ represents the scale factor, b the translation and $a, b \in \mathbf{R}$. When $a = 1$ and $b = 0$, the wavelet is called the *mother wavelet*.

The *wavelet transform* of a function $x(t) \in L^2(\mathbf{R})$ is a projection of this function on the wavelet basis $\{\psi_{a,b}\}$:

$$T(a,b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}(t)dt. \quad (3)$$

For each a, the wavelet coefficients $T(a,b)$ are signals (that depend on b) which represent the matching degree between wavelet $\psi_{a,b}(t)$ and the analyzed function $x(t)$.

The signal energy at a specific scale and position can be calculated as

$$E(a,b) = |T(a,b)|^2. \quad (4)$$

The two-dimensional wavelet energy density function is called the *scalogram*.

The CWT is a suitable tool for ECG analysis because of this time-frequency representation of the signal. With the multiscale feature of WTs, the QRS complex can be distinguished from high P or T waves, noise, baseline drift, and artifacts. The important time aspect of the non-stationary ECG signal is kept. Moreover, very efficient implementations of the algorithm exist and a low computational complexity is required, allowing real-time analysis. With the aim of a QRS detection, an appropriate mother wavelet must be chosen. It must match nicely with a QRS complex, in order to emphasizes these complexes and to filter the useless noise. For more details on the wavelet transform and on the standard wavelet functions available, the interested reader can consult (Mallat, 1999; Addison, 2005; Daubechies, 1992).

4 METHOD DESCRIPTION

4.1 General Description

The detection of R spikes is a tough task due to the complexity of the ECG signal. The aim of the algorithm introduced here is to automatically find the best subset of wavelet scales for optimal R detection. For each dataset, this subset is selected on a short training sample by a supervised learning procedure. The CWT at the selected scales is then computed on the complete dataset. Finally, R spikes are detected by a hard thresholding rule on the selected wavelet coefficients.

4.2 Training

The algorithm uses a supervised learning approach: it will use the labeled information that is provided and learn the best way to adapt to the problem. Here, the labeled information that is provided is the location of the R peaks in a training dataset.

Each dataset consists in a long-term ECG signal recording (for example 24 hours). With such long recording, the problem is that a manual extraction of the R peaks cannot be performed, as detailed in the Introduction. However, asking a specialist to annotate a small part of the signal by indicating the R peaks is perfectly feasible; this annotated part will consist in labeled segments of one minute each, taken at random locations over the entire dataset. Choosing random locations along the signal is a way to obtain a representative training set maximizing the probability to include all types of beats contained in the recordings. The CWT is then computed on the training set in a wide (therefore non restrictive) range of 50 fixed scales defined as $\{s_i\}$, $1 \leq i \leq 50$. The mother wavelet $\psi(t)$ that was used in our experiments is the mexican hat wavelet, for its similarity with the regular morphology of the QRS complex. It is defined as the second derivative of the gaussian probability density function:

$$\psi(t) = (1-t^2)e^{\frac{t^2}{2}}. \quad (5)$$

In order to select the appropriate scales among the wide range of wavelet scales, one needs a criterion. A natural criterion is the percentage of correct R peaks detection on the annotated parts of the signal using the coefficients of the wavelet transform at the trial scales in the set $\{s_i\}$. A *stepwise forward* method automatically selects the best subset $\{a_k\} \subset \{s_i\}$ of scales on the basis of the detection rate. It involves starting with an empty subset, trying out at each step the trial scales one by one and including them to the model if the detection rate is improved. The procedure stops when no scale left in $\{s_i\}$ can improve the detection rate. In addition, at each step, the scales previously selected in $\{a_k\}$ are individually challenged: if their removal does not decrease the detection rate, the scale is now useless and therefore removed from the model.

The set $\{a_k\}$ of scales coming from the selection is thus made of the scales giving the best R detection when combined together. The selected set of scales is then used for R spike detection on the complete original long-term recording. Figure 1 shows an original ECG segment and the coefficients of the first selected wavelet scale.

4.3 R Detection

The learning procedure extracted $\{a_k\} \subset \{s_i\}$, the best subset of scales on the training set. Note that the scales in the subset are not necessarily consecutive, which means that different QRS shapes can be matched at different scale levels. The CWT is computed on the whole signal at the scales $\{a_k\}$.

Because of the non-stationarity of the signal, a moving window of 5 seconds length with an overlap of one second is used to cut $T(a_k,b)$ into J parts, $1 \leq j \leq J$. For each a_k and b_j, the R spikes are detected on $T(a_k,b_j)$ by a hard thresholding rule, where index b_j scans the jth window. Let us define

$$D(a_k,b_j) = \Big(\frac{d|T(a_k,b_j)|^2}{db_j}\Big)^2. \tag{6}$$

A threshold $th(a_k,j)$ is estimated as the mean of $D(a_k,b_j)$. As $D(a_k,b_j)$ has sharp peaks in the slopes of the QRS complexes, the intervals $I(a_k,j)$ satisfying

$$D(a_k,b_j) > th(a_k,j) \tag{7}$$

belong to QRS complexes. The R spikes are then defined as the maxima of $|T(a_k,b_j)|^2$ in each of the $I(a_k,j)$ intervals. All the R spikes obtained at each scale k are then merged together.

4.4 Post-processing

A last step of post-processing makes sure that T waves or Q and S spikes have not been wrongly labeled as a R. If two or more R spikes were detected in a window smaller than 250ms (two heartbeats cannot physiologically happen in less than 250ms (Christov, 2004)), the algorithm keeps only the peak which has the highest value on the original ECG.

5 RESULTS AND VALIDATION

The learning of the model on the training set and the assessment of performances must be done using an annotated database. The public standard MIT-BIH arrhythmia database (Goldberger et al., 2000) was used in this work. It contains 48 half-hour recordings of annotated ECG with a sampling rate of 360Hz and 11-bit resolution over a 10-mV range. The recorded signals contain different wave types and only a robust algorithm can perform well on all datasets together. Some datasets include very few artifacts and clear R peaks, but others make the detection of the R spike more difficult because of (1) abnormal QRS shapes or P and T waves, (2) low signal-to-noise ratio, (3) heavy baseline drifts, (4) lots of non normal beats such as premature ventricular contraction, left bundle branch block beat, atrial premature contraction etc... Among the 48 available datasets, the four ones including paced beats were *a priori* rejected because they consist in a special case. After visual inspection of the data, datasets 207 and 208 were also rejected. The reason is that a representative training set of five times one minute would be hard to extract randomly as several minutes of these two datasets contain only non-labeled parts looking like a sinus wave.

The performances were assessed by evaluating two parameters as suggested in (Kohler et al., 2002). The sensitivity is measured as

$$\frac{TP}{TP+FN} \tag{8}$$

and the positive predictivity as

$$\frac{TP}{TP+FP}, \tag{9}$$

where TP is the number of true positive detections, FN the number of false negatives and FP the number of false positives. The error rate is also reported. It is computed by

$$\frac{FN+FP}{n_{QRS}}, \tag{10}$$

where n_{QRS} is the total number of QRS labeled in a dataset. On the database, the algorithm obtains an average sensitivity rate of 99.7% and average positive predictivity rate of 99.7%. The average error rate is below one percent. To our knowledge, only three R

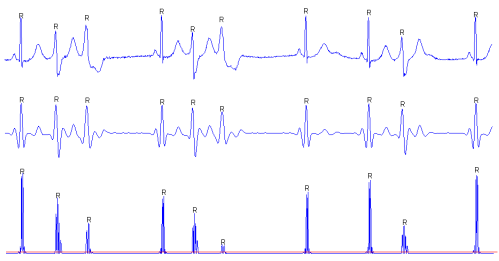


Figure 1: Example of an original ECG segment (upper plot), the first selected wavelet scale (middle plot) and its squared derivative (lower plot).

spike detectors based on WT reported in the literature obtained comparable results with a sensitivity and a positive predictivity of around 99.8% (Martinez et al., 2004; Li et al., 1995; Chen et al., 1997). Our algorithm achieves comparable performances without the need for a more advanced post-processing stage such as those used in these articles.

6 CONCLUSIONS

In this paper, a supervised learning algorithm for the automatic detection of R peaks in ECG is introduced. It uses the multiscale feature of the continuous wavelet transform (CWT) to emphasize the QRS complex over high P or T waves, noise, baseline drift and artifacts. The CWT keeps the important time aspect of the non-stationary ECG signal. Moreover, very efficient implementations of the CWT exist and a low computational complexity is required, allowing real-time analysis. This algorithm learns and propagates the annotations provided by a physician on a small annotated segment. For this purpose, the method selects the best subset of wavelet scales on a representative training set by a stepwise forward procedure. The forward procedure allows to select scales that are not necessarily consecutive and it does not a priori restrict the range of computed scales on an empirical basis. It allows a complete different set of scales to be selected for each ECG signal, based on its characteristics. The selected scales are then used on the original long-term ECG signal recording and a hard thresholding rule is applied on the derivative of the wavelet coefficients to label the R spikes. The method is robust and does not require any preprocessing stage. The selection procedure can be generalized in order to detect other ECG features such as the P and T wave.

Experiments on the public annotated MIT-BIH database lead to a sensitivity of 99.7% and a positive predictivity of 99.7% without the need of an advanced post-processing stage on the detected peaks. To our knowledge, only three R spike detectors based on WT reported in the literature obtained comparable results, while requiring a more complex post processing stage.

Further works will include: (1) The development of a more advanced thresholding rule that takes the peaks detected so far into account; (2) the use of a more advanced post-processing stage to eliminate wrong detections; (3) the design of an automatic selection of the best mother wavelet by the same learning methodology; (4) the generalization of the method for the detection of other ECG features such as P or T wave.

ACKNOWLEDGEMENTS

This work was partly supported by the Belgian “Région Wallonne” ADVENS convention 4994 project and by the Belgian “Région de Bruxelles-Capitale” BEATS project.

REFERENCES

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C., and Suri, J. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing*, Nov 17.

Addison, P. D. (2005). Wavelet transform and the ecg: a review. *Physiological Measurement*, 26:155–199.

Algra, A. and Zeelenberg, H. (1987). An algorithm for computer measurement of qt intervals in the 24 h ecg. *Proceedings of the IEEE Computer Society Press*, page 1179.

Chen, S., Chen, H., and Chan, H. (1997). Dsp implementation of wavelet transform for real time ecg waveforms detection and heart rate analysis. *Computer Methods and program in Biomedicine*, 55(1):35–44.

Chen, S., Chen, H., and Chan, H. (2006). A real-time qrs detection method based on moving-averaging incorporating with wavelet denoising. *Comput Methods Programs Biomed.*, 82(3):187–95.

Chen, Y., Yan, Z., and He, W. (2005). Detection of qrs-wave in electrocardiogram: Based on kalman and wavelets. *Conf Proc IEEE Eng Med Biol Soc.*, 3:2758–60.

Christov, I. I. (2004). Real time electrocardiogram qrs detection using combined adaptive threshold. *BioMedical Engineering OnLine*, 3(1):28.

Clavier, L., Boucher, J.-M., Lepage, R., Blanc, J.-J., and Cornily, J.-C. (2002). Automatic p-wave analysis of patients prone to atrial fibrillation. *Med Biol Eng Comp*, 40:6371.

Daubechies, I. (1992). *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*. Soc for Industrial & Applied Math.

Dobbs, S., Schmitt, N., and Ozemek, H. (1984). Qrs detection by template matching using real-time correlation on a microcomputer. *Journal of clinical engineering*, 9(3):197–212.

Fard, P. J. M., Moradi, M., and Tajvidi, M. (2007). A novel approach in r peak detection using hybrid complex wavelet (hcw). *International Journal of Cardiology*, In press(Available online 27 March 2007).

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C.-K., and Stanley, H. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.

Kadambe, S., Murray, R., and Boudreaux-Bartels, G. (1999). Wavelet transform-based qrs complex detector. *IEEE Transactions on Biomedical Engineering*, 46:838–48.

Koeleman, A., Ros, H., and van den Akker, T. (1985). Beat-to-beat interval measurement in the electrocardiogram. *Med Biol Eng Comp*, 23:2139.

Kohler, B., Hennig, C., and Orglmeister, R. (2002). The principles of software qrs detection. *IEEE Eng Med Biol Mag.*, 2(1):42–57.

Li, C., Zheng, C., and Tai, C. (1995). Detection of ecg characteristic points using wavelet transform. *IEEE Trans.Biomed*, 42(1):21–28.

Madeiro, J., Cortez, P., Oliveira, F., and Siqueira, R. (2007). A new approach to qrs segmentation based on wavelet bases and adaptive threshold technique. *Medical Engineering and Physics*, 29:2637.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis and Its Applications)*. IEEE press, San Diego. ISBN 978-0124666061.

Martinez, J., Almeida, R., Olmos, S., Rocha, A., and Laguna, P. (2004). A wavelet-based ecg delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51:570–81.

Nygards, M. and Sornmo, L. (1983). Delineation of the qrs complex using the envelope of the ecg. *Med Biol Eng Comp*, 21:53847.

Pahlm, O. and Sornmo, L. (1984). Software qrs detection in ambulatory monitoringa review. *Med Biol Eng Comp*, 22:28997.

Sahambi, J., Tandon, S., and Bhatt, R. (2000). An automated approach to beat-by-beat qt-interval analysis. *IEEE Eng. Med. Biol. Mag.*, 19(3):97–101.

Senhadji, L., Carrault, G., Bellanger, J., and Passariello, G. (1995). Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Eng. Med. Biol. Mag.*, 149(2):167–173.

Shyuand, L.-Y., Wu, Y.-H., and Hu, W. (2004). Using wavelet transform and fuzzy neural network for vpc detection from the holter ecg. *IEEE Transactions on Biomedical Engineering*, 51:1269–73.

Surez, K. ., Silva, J., Berthoumieu, Y., Gomis, P., and Najim, M. (2007). Ecg beat detection using a geometrical matching approach. *IEEE Transactions on Biomedical Engineering*, 54(4):641–50.

Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology (1996). Heart-rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–65.

Vullings, H., Verhaegen, M., and Verbruggen, H. (1998). Automated ecg segmentation with dynamic time warping. *Proceedings of the 20th Annual International Conference on IEEE Engineering in Medicine and Biology Society*, page 1636.

Xue, X., Hu, Y., and Tompkins, W. (1992). Neural-network-based adaptive matched filtering for qrs detection. *IEEE Transactions on Biomedical Engineering*, 32(4):317–329.

REGISTRATION AND RETRIEVAL OF ELONGATED STRUCTURES IN MEDICAL IMAGES

Alexei Manso Correa Machado and Christiano Augusto Caldas Teixeira
Pontifical Catholic University of Minas Gerais, Av. Dom Jose Gaspar, 500, Belo Horizonte, MG, Brazil
alexei@pucminas.br, chriteixeira@gmail.com

Keywords: Medical imaging, morphology, image registration, information retrieval.

Abstract: This work aims at proposing a set of methods to describe, register and retrieve images of elongated structures from a database based on their shape content. We propose a registration algorithm that jointly takes into account the gross shape of the structure and the shape of its boundary, resulting in anatomically consistent deformations. The method determines a medial axis that represents the full extent of the structure with no branches. Registration follows the linear elasticity model and is implemented through dynamic programming. Discriminative anatomic features are computed from the results of registration and used as variables in a content-based image retrieval system. A case study on the morphology of the corpus callosum in the chromosome 22q11.2 deletion syndrome illustrates the effectiveness of the method and corroborates the hypothesis that retrieval systems may also act as knowledge discovery tools.

1 INTRODUCTION

Elongated structures such as vessels, bones and brain ventricles are of interest in many problems and applications (Toledo et al., 2000; Staal, 2004). Those structures have in common the fact that their gross shape can be efficiently represented by centerlines or medial axes. Contour may present important anatomical features, but the overall shape is, if not more, as important as the shape of the boundary.

This work aims at proposing a set of methods to describe, register and ultimately retrieve images of elongated structures from a database based on their shape content. Image registration techniques have been widely used in morphometry, as it provides detailed description of the anatomy, taking a reference image as a basis for comparison. Registration algorithms are nevertheless computationally costly and, when applied to the whole image or to the boundary of elongated structures, may yield unsatisfactory results. A contribution of this work is a registration algorithm that takes into account both the gross shape of the structure and the shape of its boundary, with emphasis to the former aspect.

Figure 1 shows a schematic of a content-based image retrieval (CBIR) system that follows this approach. A set of images depicting elongated structures is segmented and the structures represented by their boundaries and medial axes. Another image, taken as a common reference, is deformed through elastic registration so as to align its anatomy with the anatomy of the images in the dataset. The result of registration is a mapping function from each point in the reference to a point in the target image that enable detailed shape description. After the structures have been described, e.g. based on the curvature of their boundaries and medial axes, they are stored in the database for future searching. The querying phase follows the same steps used to convert the images into descriptive features. The query image converted to the corresponding feature vector is compared with the database, the most similar images are retrieved and presented to the user. The user may rank the results according to their relevance, choose one of the retrieved images as a new query or redefine a region of interest that should be given greater priority in the next retrieving iteration. The query vector is therefore updated taking into account the user's feedback.

The characterization of the gross shape is critical to the registration and retrieval of elongated structures. We also present a semi-automatic solution to the extraction of a medial axis that represents the full extent of the structure with no branches. Finally, discriminative anatomic features are computed from the results of registration and used as variables in a CBIR system. A case study on the morphology of the corpus

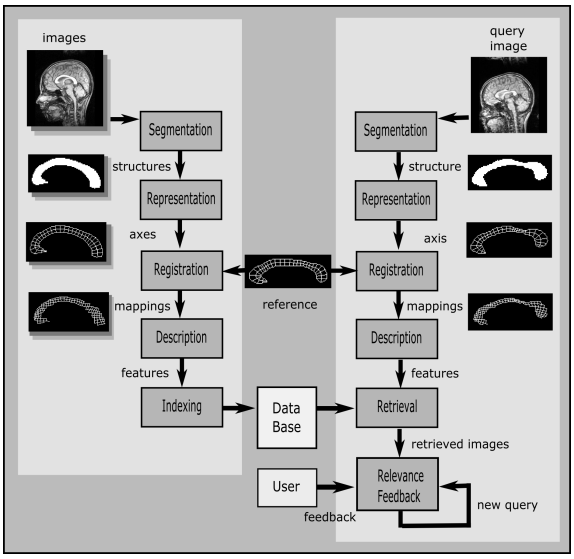


Figure 1: Schematic of a CBIR system based on registration. The left part of the scheme shows the steps performed off-line for each image in the database. The on-line part of the retrieval process is shown in the right. The link between the on-line and off-line phases is the reference image that is registered to the query and to the database, establishing a basis for shape comparison.

callosum in the chromosome 22q11.2 deletion syndrome illustrates the effectiveness of the method and corroborates the hypothesis that CBIR systems may also act as knowledge discovery tools.

2 RELATED WORKS

The representation of elongated structures through single sequences of connected points that describe their intrinsic geometry has been extensively studied. Pioneered by Blum and Nagel (Blum and Nagel, 1978), the use of medial axes to describe 2D shapes is based on the removal of points in the boundary until the gross shape is minimally represented. Many skeleton and thinning algorithms can be found in the literature, revealing the difficulty on determining a standard definition for medial axis (Dvies and Plummer, 1981). Other more complex models include the medial representations (Pizer et al., 2003; Yushkevich et al., 2003), in which the medial axis and a radial scalar field are parametrically described such that the boundary can be further reconstructed, and the medial profiles (Hamarneh et al., 2004), that provide a shape representation and deformation operators that can be used to derive shape distributions.

Registration is considered one of the most important approaches to provide detailed description of shape. Automatic registration algorithms (McInerney and Terzopoulos, 1996; Toga, 1999) may be applied to the contour (Cootes et al., 1994; Davatzikos and Prince, 1995) or medial axis (Pizer et al., 1996; Golland et al., 1999) of specific structures. Registration is also used together with the medial axis transform (Xie and Heng, 2005) to align the anatomy of structures based on their skeletons.

Retrieval of images based on their content is still in its infancy. Smeulders (Smeulders et al., 2000) and Lew (Lew et al., 2006) present comprehensive discussions on the main aspects and challenges of image retrieval. Muller (Muller et al., 2004) shows how CBIR systems can be used to retrieve images in general medical databases. In the next section, we discuss the specific issues related to the registration and retrieval of images depicting elongated structures and propose a registration algorithm that jointly considers the axis and boundaries of such structures.

3 METHODS

The proposed image retrieval method can be divided into four steps: midline extraction, registration, description and retrieval.

3.1 Midline Extraction

A midline can be defined as a curve that splits the structure into dorsal and ventral regions, such that, at any point, the perpendicular line segments connecting the midline to dorsal and ventral parts of the boundary have roughly the same length (properties of perpendicularity and congruency). Midline extraction starts by determining a skeleton based on a variation of the thinning algorithm described by Gonzalez and Woods (Gonzalez and Woods, 2002), for 8-connected objects. Object points are labeled as 1 and the background is set to 0. In order for the curve to fully extend from one extremity to the other, two object points are manually chosen and forced to be respectively the starting and ending points of the skeleton.Additionally, the thinning algorithm is modified so as to prune any other branches of the structure's skeleton. The final curve is, therefore, a single sequence of pixels, each one connected to two neighbors, with the exception of the starting and ending points.

The following algorithm summarizes skeleton extraction, where p_1 and p_2 are the endpoints; the neighbors of p are denoted as n_i, numbered counterclockwise from 0 (east) to 7 (southeast); function N returns the number of neighbors of p that belong to the object, i.e., $N(p) = \sum_i n_i$; and function S returns the

number of connected sequences of object points in the neighborhood of p, i.e., read as an 8-bit string, the neighbors of p must match the regular expression $0^+1^+0^* \bigcup 1^+0^+1^*$. It can be shown that only 42 neighborhood configurations satisfy the condition to mark a point, so that the algorithm can be efficiently implemented using look-up tables:

Repeat
  For each point p of the object, $p \notin \{p_1, p_2\}$, do
    If $N(p) < 7$ and $S(p) = 1$ and $n_0 n_6 (n_2 + n_4) = 0$
      Mark p to be removed;
  Remove marked points;
  For each point p of the object, $p \notin \{p_1, p_2\}$, do
    If $N(p) < 7$ and $S(p) = 1$ and $n_2 n_4 (n_0 + n_6) = 0$
      Mark p to be removed;
  Remove marked points;
until no more points can be removed.

The linear length of the skeleton is computed considering the distances between each pair of consecutive pixels: pixels connected by a face with distance equals to 1 and the ones connected by a vertex with distance equals to $\sqrt{2}$. The coordinates of the pixels are smoothed and interpolated so as to yield an isotropic rotation-invariant representation of the midline. The derivative of this curve, taken at equidistant points, guides the computation of perpendicular segments that link the dorsal and ventral boundaries of the structure. Problems may occur in regions where the midline presents increased curvature. In this case, it may be impossible to satisfy the requirements of perpendicularity and congruency for the segments. Figure 2 shows an example where two consecutive segments intersect each other as the result of increased midline curvature. A solution for this problem is to violate the property of perpendicularity so that points with increasing coordinates at the midline will be connected to points of non-decreasing coordinates at both boundaries. It is however expected that elongated structures will not frequently incur in this problem.

The curvature (second derivative) of the midline can be determined based on the k-curvature metric, that is defined in each point $p_i = (x_i, y_i)$ as the difference between the average of the derivatives at the k next points and the average of the derivatives at the k previous points (including p_i):

$$kcurv(p_i) = \frac{1}{k}\left(\sum_{j=i+1}^{i+k} d(p_j) - \sum_{j=i-k+1}^{i} d(p_j)\right), \quad (1)$$

$$d(p_j) = \tan^{-1}(x_j - x_{j-1}, y_j - y_{j-1}).$$

Parameter k should be empirically chosen so as to provide enough smoothness. The midline curve should be extrapolated at the extremities (e.g. based on autoregression), so that the curvatures will be computed over all the midline extension. Analogously, the curvature at the dorsal and ventral boundaries should be computed at the intersection of the segments. The curvatures at the midline and boundaries will play a fundamental role as a measure of similarity during registration.

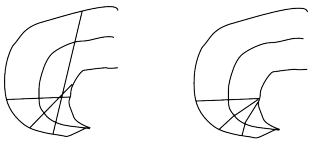


Figure 2: Example where consecutive segments intersect each other as the result of increased midline curvature (a) and the solution to the problem (b).

3.2 Image Registration

The images in the database should be registered to a reference in order to establish a common basis for comparison. Image registration can be stated as the process of determining a correspondence between each point p in the midline of the reference image to a point $u(p)$ in the midline of the subject image. Let $C_M(p) = kcurv(p) - kcurv(u(p))$ be the difference between the k-curvature taken at point p in the reference midline and the k-curvature taken at point $u(p)$ in the subject midline. Analogously, let C_D and C_V be the same difference function computed respectively at the intersection points of the perpendicular segments emanating from the midline with the dorsal and ventral boundaries.

The cost function to be minimized is given as

$$cost = D - S, \quad (2)$$

where D is the deformation penalty and S is the similarity between the curvatures of registered points of the midline, dorsal and ventral boundaries, given as

$$D = \alpha \int_0^1 \left(\frac{du(p)}{dp}\right)^2 dp + \beta \int_0^1 \left(\frac{d^2u(p)}{dp^2}\right)^2 dp,$$

$$S = \sum_{i \in \{M,D,V\}} \gamma_i \int_0^1 C_i(p)^2 dp \quad (3)$$

Parameters α and β weight the amount and smoothness of deformation, respectively. Parameters γ_M, γ_D and γ_V are negative and weight the importance of the similarity terms computed respectively for the midline, dorsal and ventral boundaries.

Registration is performed through dynamic programming, in which equidistant points in the reference midline are mapped to points in the midlines of the database by minimizing the cost function in (2). After registering the midlines and corresponding

boundaries, thin plate splines (Barrodale et al., 1993) are used to interpolate the warping applied to these curves to the whole structure, so that each pixel in the reference image is assigned a displacement vector.

An advantage of the proposed registration algorithm is that it will always map a segment perpendicular to the reference midline to a segment perpendicular to the midline of the subject image. This is a very important constraint to be observed when dealing with elongated structures. Fig 3 shows two examples where an image registration algorithm based only on the boundary or only on the midline would fail to provide satisfactory deformations. The structure in (a) is the reference, whose boundary points A and B must be found correspondence in the other structures. A registration algorithm that takes into account only the boundaries would map point A to C (correctly), but B to E instead of D, since the boundary curvature in B is more similar to the curvature in E than it is in D. If, on the other hand, the algorithm is based only on the curvature of the midline, the registration of the reference to the structure in (c) would probably map the segment $\overline{AB}$ to $\overline{HI}$ instead of $\overline{FG}$, ignoring the similarity between the curvatures at the boundaries. The similarity function proposed in (3) avoid both mistakes, since the curvatures at the midline, dorsal and ventral boundaries are jointly taken into account.

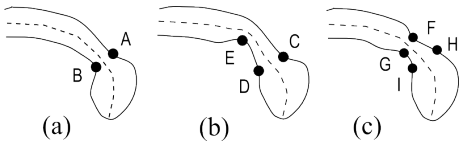


Figure 3: Examples of unsatisfactory registration of the segment $\overline{AB}$ in the reference structure (a) to segment $\overline{CE}$ in (b) and to segment $\overline{HI}$ in (c). An algorithm based on both the boundary and midline would correctly map $\overline{AB}$ to $\overline{CD}$ and $\overline{FG}$.

Evaluating the effectiveness of registration methods is always a difficult task, as ground truth data is usually inexistent, particularly when the structure being registered does not present well-defined landmarks. Alternatively, landmarks may be chosen by experts, but in this case human subjectivity and lack of repeatability should be considered in the analysis. In this work, we designed an interactive interface in which an expert chooses a set of landmarks in the reference structure and the corresponding loci in the subjects. The procedure is repeated after 2 weeks, in order to evaluate repeatability. The results achieved by automatic registration are compared to the mapping provided by the expert: if the result falls within the interval of values provided by the expert, it is considered satisfactory, otherwise the distance in millimeters to nearest value is stored and averaged.

3.3 Description

The output of registration is a displacement field that maps each pixel of the reference image to a point in the subject. From this set of vectors, it is possible to obtain diverse measurements that describe the imaged objects, such as point-wise area and length variation, curvature of axes and contours, relationships between axes of orientation, moments and other shape descriptors. Feature selection is a fundamental step in image retrieval systems, as it determines the effectiveness and efficiency of many algorithms. The set of features that will represent the objects should be concise and discriminative, as distinguishing features facilitates the retrieval of relevant images, while non-relevant characteristics are confounders. Feature selection and information retrieval are synergetic steps: while the choice of distinguishing features increases the relevance of retrieval results, retrieval itself act as a "mining" tool, selecting the features that discriminate between classes of images. This is the fundamental relationship that characterizes image retrieval as a potential knowledge discovery methodology. In this work, objects were described as vectors of k-curvatures (1) taken at each matched point of the subjects, after being registered to the reference.

3.4 Image Retrieval

In a CBIR system, the user presents an image as a query, which is registered to the reference image. The features obtained from the resulting mapping function are compared to the features of the images in the database, which have been previously processed and registered to the same reference. Following a measure of similarity, the most similar images are retrieved and presented to the user.

The model used to determine the similarity between two images was the Euclidean distance (Del Bimbo, 1999). If $\mathbf{q}$ is the feature vector representing the query and $\mathbf{v}_k$ is the feature vector representation of image k in the database, the similarity between them can be computed as

$$sim(\mathbf{v}_k, \mathbf{q}) = ((\mathbf{v}_k - \mathbf{q})^T (\mathbf{v}_k - \mathbf{q}))^{1/2}$$

The performance of an image retrieval system can be evaluated by computing two metrics (Del Bimbo, 1999): The recall of the system is the ability to retrieve relevant images. It is defined as the ratio between the number of retrieved images considered relevant and the total number of relevant images in the database. The precision reflects the ability of the system to retrieve only relevant images. It is defined as

the ratio between the number of retrieved images considered relevant and the total number of retrieved images. The plot of recall $\times$ precision gives an estimate of the overall effectiveness of a CBIR system, as a compromise between both performance metrics is expected.

4 EXPERIMENTS

We illustrate the proposed registration-based retrieval system with a case study on the morphology of the corpus callosum in the chromosome 22q11.2 deletion syndrome (DS22q11.2). The DS22q11.2 is an example of genetic abnormality for which many hypotheses on anatomical differences have been recently stated (Machado et al., 2007). This syndrome is the result of a 1.5 - 3Mb microdeletion on the long arm of chromosome 22 and is characterized by a range of medical manifestations that include cardiac, palatal and immune disorders, as well as particular problems in cognitive domains associated with the orienting and executive attention systems and with numerically related processing. Recent studies have drawn particular attention to changes in the corpus callosum — the largest bundle of axons connecting the two hemispheres of the brain, as differences in the shape of this structure may indicate changes in brain connectivity that may be related to the observed cognitive impairments (Simon et al., 2005). We hypothesized that an image retrieval system would be able to retrieve images of subjects sharing the same diagnosis, based on a shape representation of the corpus callosum, if the features used to index the images could be considered discriminative for the syndrome. In this sense, the system would reveal the most distinguishing features associated with the disease.

Participants in this study were 18 children with chromosome 22q11.2 deletion syndrome, ranging in age from 7.3 to 14.0 years (mean,S.D.=9.9,1.4 years) and 18 typically developing control children, ranging in age from 7.5 to 14.2 years (mean,S.D.=10.4,2.0 years) (Simon et al., 2005). Magnetic resonance imaging was performed on a 1.5 Tesla Siemens MAGNETOM Vision scanner (Siemens Medical Solutions, Erlangen, Germany). For each subject, a high-resolution three-dimensional structural MRI was obtained using a T1-weighted magnetization prepared rapid gradient echo (MP-RAGE) sequence with the following parameters: repetition time (TR) = 9.7 ms, echo time (TE) = 4 ms, flip angle = 12(, number of excitations = 1, matrix size = 256x256, slice thickness = 1.0 mm, 160 sagittal slices, in-plane resolution = 1x1 mm. The midsagittal slice of each brain image volume was manually extracted as the best plane spanning the interhemispheric fissure, and on which the anterior and posterior commissures and the cerebral aqueduct were visible.

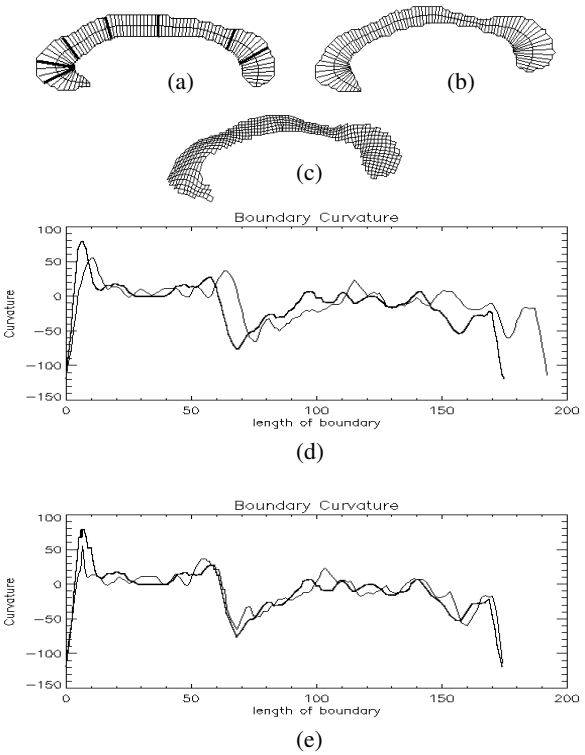


Figure 4: An example of registration. The midline and boundary of the reference (a) is registered to the subject (b) and the result interpolated to the whole structure (c). The original plot of the boundary curvatures (d) and result of registration (e) are also shown, where the curvatures of the template and subject are represented by thick and thin lines, respectively. The 7 landmarks used for registration evaluation, numbered from left to right, are depicted in (a) with thick lines.

The callosa in the midsagittal images were segmented by manual thresholding and delineation. The boundaries of the callosa were automatically determined using the Rosenfeld algorithm for 8-connected contours (Gonzalez and Woods, 2002). The midlines of the callosa were also extracted based on the algorithm proposed in Section 3.1 and interpolated so as to yield an isotropic rotation-invariant representation, in which any two consecutive sampled points were 1 mm apart. The pointwise curvature of the callosum midline was computed for each subject, using the k-curvature metric (1), where k was empirically chosen to be 10% of the length of the midline, so as to provide enough smoothness.

Shape measurement was performed, by aligning a reference image of the callosum to subject callosa. One of the control subjects was arbitrarily chosen as

the reference. The midline of the reference, sampled at 87 equidistant points, was registered to the subjects' midlines based on the cost function described in (2) with parameters α=0.001, β=1000.0 and γ_i=-1.0 mm^2/degree^2 for $i \in \{M, D, V\}$, which were empirically determined. The midline curves of the subject callosa were interpolated to provide sub-pixel precision (0.5 mm). The result of registration was a mapping from each of the 87 points in the reference to corresponding points in the subjects. Registration took 7.78 seconds to compute. All methods were implemented in IDL language (Research Systems) and run in a 1.1 GHz Intel Celeron processor computer with 256 MB of RAM, under Windows XP operating system.

Figure 4 shows an example of registration where the reference image described through its midline and perpendicular segments (a) is deformed to match the subject (b). The resulting deformation is shown as a warped grid (c). A plot of the original k-curvatures (in degrees/mm) at the boundaries of both images (in mm), taken counterclockwise from the leftmost endpoint of the midline, is given in (d) and the resulting registration is depicted in (e). The effectiveness of registration was evaluated based on 2 sets of landmarks provided by an expert, taken in an interval of 2 weeks. Seven landmarks were defined at the reference, from anterior to posterior callosum (Figure 4a), and the expert was asked to determine their corresponding loci at each of the 36 subjects. The set of 504 landmarks were compared to the results of registration. Table 1 summarizes the results, where it is possible to compare the average error of the method with the variability of measures provided by the expert, for each landmark. The average error of the method for the whole set of landmarks was 1.7 mm, a satisfactory result considering that the average variability of the expert's measures was 1.2 mm. Larger errors were observed at landmarks 3 and 4 (callosal body) where the subjects present larger variability with respect to curvature. The best results were achieved at landmarks 5 and 6 (posterior callosum) where the errors obtained with automatic registration were smaller than the average variability observed in manual registration.

The results of image retrieval were evaluated with the aid of a simple retrieval environment. Initially, the user browses the database and chooses an image that will represent the query. The system ranks the remaining images, showing the n most relevant to the user appraisal. In this study, we considered as relevant the images that shared the same diagnosis of the query (with or without the deletion). Following the recent findings on anatomic differences in the callosum of these populations (Machado et al., 2007)(see Figure 5), an effective CBIR should be able to retrieve images sharing the same diagnosis, unless outliers would be present in the database.

(a) (b)

Figure 5: Mean callosal shape for the typically developing children (a) and children with the deletion (b). Controls have shorter, more curved anterior callosum (rostrum and genu) and less curved midbody. Children with the deletion present more arched callosum (larger height/length ratio).

Table 1: Average error (mm) for each landmark, considering manual and automatic registration.

Landmark	0	1	2	3	4	5	6
Manual	0.6	0.4	0.7	1.7	1.3	2.6	1.0
Automatic	1.0	0.9	0.8	3.7	2.4	1.6	0.9

An example of the results of image retrieval is shown in Figure 6. The query image presented by the user (a) is registered to the same reference used in the registration of the images stored in the database. The 10 images that yield greater similarity with respect to the curvature of the midline and boundary are retrieved and displayed (b). Images of controls are shown in gray and images of children with the deletion are shown in black. A plot of the recall $\times$ precision computed after the retrieval of each of the 17 relevant images in the database is presented in (d). In this case, the query is a typical control, yielding high precision.

An example in which an outlier is retrieved is given in Figure 7. The third retrieved image is a control with arched callosum, whereas the query is a child with the deletion. In this case, the precision is affected. Worse result occurs when the query itself is an outlier, as exampled in Figure 8. In this case, the query is a control with longer, less curved rostrum (left-most end of the midline) that is more common in children with the deletion. As a consequence, the precision is drastically affected, staying bellow 50% from the second retrieved image, a level that would be expected by pure chance.

5 CONCLUSIONS

We have addressed the problem of registering and retrieving images of elongated structures. Traditional

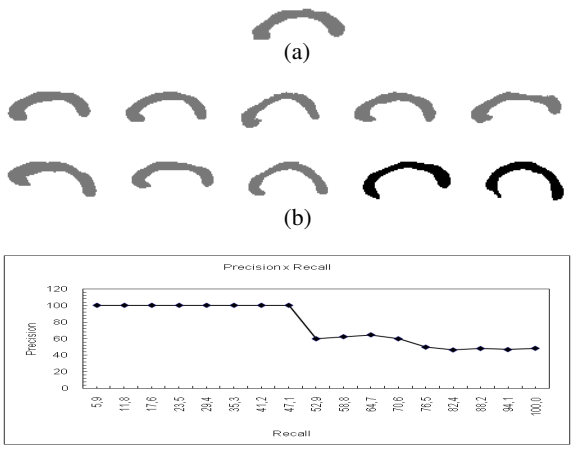


Figure 6: Example of a query image (a) and the result of retrieval (b). The plot of recall × precision is shown in (c).

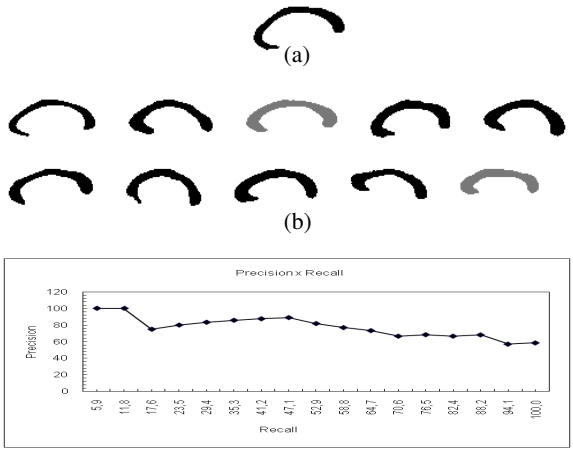


Figure 7: Example of a query image (a) and the result of retrieval (b). In this case, the third best-ranked image is an outlier. The plot of recall × precision is shown in (c).

registration methods may yield anatomically inconsistent results while applying warping models only to the structure's contour or medial axis. The method proposed in this paper jointly registers the medial axis, dorsal and ventral boundaries, avoiding distortions that may impact substantially in the results of further morphometric analyses, hypothesis testing or image retrieval.

The method deserves more systematic evaluation procedures, as visual inspection is subjective and difficult to quantify. A case study on the morphology of the corpus callosum in the 22q11.2 deletion syndrome was used to illustrate the ability of registration to provide effective image retrieval. In the experiments, diagnosis was considered as the ground truth to evaluate the performance of the retrieval system. Although evidences of shape differences between controls and children with the deletion exist, outliers make evaluation a difficult task. A deficiency of the method is the requirement for manual choice of the midline endpoints, so a fully automated algorithm is already being designed. Another well-known disadvantage of registration-driven retrieval methods is its inadequacy to indexing, limiting the application of these systems to small datasets. Furthermore, the vector model that exhibits excellent performance in text retrieval is not a consensus when dealing with images.

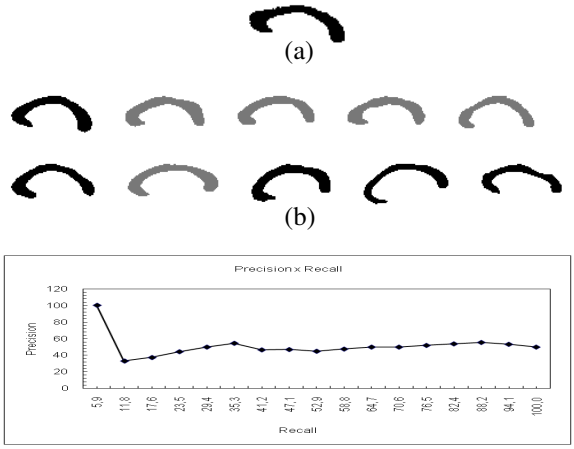


Figure 8: Example of a query image (a) and the result of retrieval (b). In this case, the query is an outlier, yielding poor performance (c).

Relevance feedback is an important step that deserves attention. Different similarity functions and query updating models may enhance the effectiveness of image retrieval, as the user's preferences are more rapidly met. Experiments have shown that when the set of features is restricted to specific regions of interest, the precision is enhanced. In the case of the study on the corpus callosum morphometry, restricting the computation of similarity to the anterior-most part of the structure, where the differences between groups are more evident, has increased the number of retrieved images that share the same diagnosis. This ability to cluster images of the same group may qualify image retrieval as a potential knowledge discovery tool. It implements new levels of supporting environments and opens new perspectives to exploratory research in image databases.

ACKNOWLEDGEMENTS

This work was partially supported by FAPEMIG, PUC Minas and CNPq grant 20043054198. The authors are grateful to the University of Pennsylvania for sharing the corpus callosum data.

REFERENCES

Barrodale, I., Skea, D., Berkley, M., Kuwahara, R., and Poeckert, R. (1993). Warping digital images using thin plate splines. *Pattern Recognition*, 26(2):375–376.

Blum, H. and Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, 10(3):167–180.

Cootes, T., Hill, A., Taylor, C. J., and Haslam, J. (1994). Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365.

Davatzikos, C. and Prince, J. (1995). An active contour model for mapping the cortex. *IEEE Transactions on Medical Imaging*, 14:65–80.

Del Bimbo, A. (1999). *Visual Information Retrieval*. Morgan Kaufman.

Dvies, E. R. and Plummer, A. P. (1981). Thinning algorithms: a critique and a new methodology. *Pattern Recognition*, 14:53–63.

Golland, P., Grimson, W. E. L., and Kikinis, R. (1999). Statistical shape analysis using fixed topology skeletons: Corpus callosum study. *Lecture Notes in Computer Science*, 1613:382–387.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Prentice-Hall, Upper Saddle River.

Hamarneh, G., Abu-Gharbieh, R., and McInerney, T. (2004). Medial profiles for modeling deformation and statistical analysis of shape and their use in medical image segmentation. *International Journal of Shape Modeling*, 10(2):187–209.

Lew, M., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19.

Machado, A., Simon, T., Nguyen, V., McDonald-McGinn, D., Zackai, E., and Gee, J. (2007). Corpus callosum morphology and ventricular size in chromosome 22q11.2 deletion syndrome. *Brain Research*, 1131:197–210.

McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108.

Muller, H., Michoux, N., Bandon, D., and Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23.

Pizer, S. M., Fletcher, P. T., Joshi, S., Thall, A., Chen, J. Z., Fritsch, D. S., Gash, A. G., Glotzer, J. M., Jiroutek, M. R., Lu, C., Muller, K. E., Tracton, G., Yushkevich, and Chaney, E. (2003). Deformable m-reps for 3D medical image segmentation. *International Journal of Computer Vision*, 55(2):85–106.

Pizer, S. M., Fritsch, D. S., Yushkevich, P., Johnson, V., and Chaney, E. (1996). Segmentation, registration and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18(10):851–865.

Simon, T. J., Ding, L., Bish, J. P., McDonald-McGinn, D. M., Zackai, E. H., and Gee, J. C. (2005). Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. *Neuroimage*, 25:169–180.

Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380.

Staal, J. (2004). *Segmentation of elongated structures in medical imaging*. PrintPartners Ipskamp, Enschede.

Toga, A. W. (1999). *Brain Warping*. Academic Press, New York.

Toledo, R., Orriols, X., Binefa, X., Radeva, P., Vitria, J., and Villanueva, J. (2000). Tracking of elongated structures using statistical snakes. In *Proceedings of the CVPR*, pages 157–162, Hilton Head Island.

Xie, J. and Heng, P. A. (2005). Shape modeling using automatic landmarking. *Lecture Notes in Computer Science*, 3750:709–716.

Yushkevich, P., Fletcher, P. T., Joshi, S., Thall, A., and Pizer, S. (2003). Continuous medial representations for geometric object modeling in 2d and 3d. *Image and Vision Computing*, 21(1):17–28.

NETWORK TOMOGRAPHY-BASED TRACKING FOR INTRACELLULAR TRAFFIC ANALYSIS IN FLUORESCENCE MICROSCOPY IMAGING

Thierry Pécot[1,2], Charles Kervrann[1,2] and Patrick Bouthemy[1]
[1]*IRISA/INRIA, Campus universitaire de Beaulieu, F-35042 Rennes, France*
[2]*INRA, UR341 Mathématiques et Informatique Appliquées, F-78352 Jouy-en-Josas*
thpecot@irisa.fr, ckervran@irisa.fr, bouthemy@irisa.fr

Keywords: Object tracking, fluorescence microscopy, network tomography, Voronoi diagram, trafficking, membrane transport.

Abstract: Determination of the sub-cellular localization and dynamics of any proteins is an important step towards the understanding of multi-molecular complexes in a cellular context. Green Fluorescent Protein (GFP)-tagging and time-lapse fluorescence microscopy allows to acquire multidimensional data on rapid cellular activities, and then make possible the analysis of proteins of interest. Consequently, novel techniques of image analysis are needed to quantify dynamics of biological processes observed in such image sequences. In biological trafficking analysis, the previous tracking methods do not manage when many small and poorly distinguishable objects interact. Nevertheless, an another way of tracking that usually consists in determining the full trajectories of all the objects, can be more relevant. General information about the traffic like the regions of origin and destination of the moving objects represent interesting features for analysis. In this paper, we propose to estimate the paths (regions of origin and destination) used by the objects of interest, and the proportions of moving objects for each path. This can be accomplished by exploiting the recent advances in Network Tomography (NT) commonly used in network communications. This idea is demonstrated on real image sequences for the Rab6 protein, a GTPase involved in the regulation of intracellular membrane trafficking.

1 INTRODUCTION

Small GTPases play a key role in many aspects of cell biology: control of cell growth and differentiation, regulation of cell adhesion and cell movement, organization of the actin cytoskeleton, and regulation of intracellular vesicular transport. The small GTPases Rab proteins are important regulators of trafficking within the membrane. Each member of this family (60 described in human cells) exists under different dynamic states in the cell: *i*) diffusion in the cytosol; *ii*) exchanges between the cytosol and the membranes; *iii*) vesicular transport. The Rab protein family plays an essential role in the dynamics of the transport vesicles and their targeting/anchoring with the acceptor membranes. Studying the role of Rab proteins inside multiprotein complexes is then fundamental to deeply understand the molecular mechanisms responsible for membrane transport and for the maintenance of the integrity and global architecture of the cell, in space and time.

Rab6 is located on the Golgi Apparatus membranes and the trans-Golgi network membranes. It is involved in a retrograde transport from the Golgi Apparatus to the Endoplasmic Reticulum. When Rab6 proteins embedded into vesicles are marked with GFP (Green Fluorescence Protein), they appear on the image sequence as blobs heterogeneously moving along the microtubule network. The study of the membrane trafficking by measuring the activity of small transport vesicles from donor to acceptor compartments within the cell thanks to image analysis techniques is challenging.

Rab6 trafficking is really hard to analyse as it is composed of several hundreds similar objects that are moving with variable velocities. The most commonly used tracking concept is the connexionist approach (Anderson et al., 1992; Sbalzarini and Koumoutsakos, 2005; Racine et al., 2006) consisting in detecting particles independently in each frame in a first time, and then linking the detected objects over time. But, measurements from clutter and multiple objects make the data association problem very hard to compute. From now, data association even combined with sophisticated particle filtering techniques (Smal et al., 2007) or graph-theory based methods

(Thomann et al., 2003) are problematic to track several hundreds of similar objects with a high reliability.

Deterministic approaches have also been explored. (Sibarita et al., 2006) exploits the fact that vesicles are moving along the microtubule network, and thus follow the same paths. Kymograms are used for analysing the time intensity profile of the given paths. The main limitation of the kymogram-based method is that each path is independently supervised. Another line of work consists in detecting changes in the temporal signal for a set of pixels (Bechar and Trubuil, 2006). By grouping similar temporal profiles, dynamics of vesicles can be better described.

In this paper, we propose to get around the difficult problem of data association by using an original statistical approach. The aim is to apply the Network Tomography (NT) concept to real image sequences, which is challenging for several reasons described below. Accordingly, we need to construct a graph and to propose a method to measure the activity on edges, according to the NT approach (Vardi, 1996). This is the main contribution of this paper. The NT-based approach, already applied in video surveillance (Santini, 2000; Boyd et al., 1999), allows us to track objects but only requires the detection of the objects when they move from one region to another. The estimated variables give only a general aspect of the whole traffic, but the data association, usually complex, is not needed. In this paper, we propose to adapt this NT concept to the estimation of trajectories of vesicles since it can be motivated by biological analyses. The number of vesicles that pass through each transition of the graph is estimated by solving an underconstrained optimization problem. We will demonstrate that this method is suited for understanding membrane transport. The paper is organized as follows: in Section 2, we propose to partition the image into regions of interest, and we estimate the number of moving vesicles on edges at each time step. Then, this estimation is tested on simulations. In Section 3, we estimate the regions of origin and destination for the vesicles, and these estimations are tested on a real image sequence in Section 4. Finally, we present a conclusion and the perspectives in Section 5.

2 MEASUREMENTS ON EDGES

In (Pécot et al., 2007), the idea was to extract the microtubule network, and to determine the origin and destination regions for the vesicles, and the crossings of different microtubules, all labeled as vertices in the graph. Vertices and edges (links between vertices) define the graph $\mathcal{G} = \{E, V\}$, and the activity measurements on edges correspond to the observations required to apply the NT-based approach, which amounts to estimating the origin-destination (OD) pairs for the vesicles. In other words, our goal is to determine the different paths used by the vesicles from the donor compartment to the acceptor compartment, and the proportions of vesicles for each path. However, the extraction of the microtubule network is really hard to compute, since very complex with limited spatial resolution. So we prefer to partition the image into regions and to represent the relationships between regions using a graph.

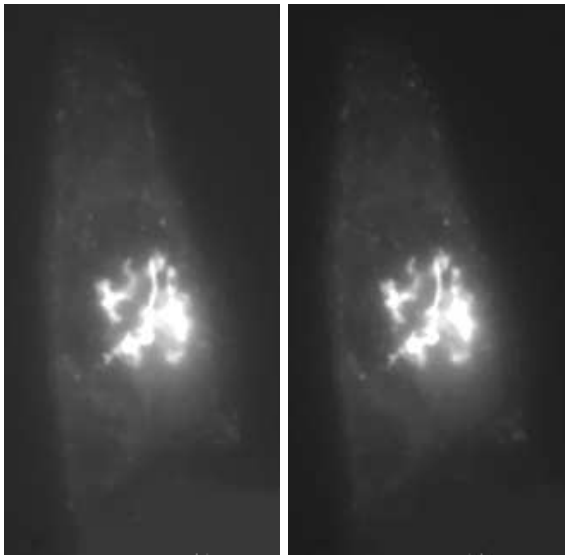


Figure 1: Images extracted from a microscopic sequence using a fast 4D deconvolution (wide-field) process at two time steps.

2.1 Image Partitionning

The Maximum Intensity Projection (MIP) map in the direction of time axis is a precious key for the partitionning of a cell compartment. Indeed, the likely regions of origin or destination appear as brighter spots in the MIP map because vesicles are temporally stocked in these areas. For illustration, the MIP map extracted from the image sequence shown in Fig. 1 is given in Fig. 2. It is established that the Golgi Apparatus is the main origin region for Rab6 protein. This region appears as a very bright region in the MIP map as shown in Fig. 2. A possible image partitionning consists in dividing the image into Voronoi cells as in (Boyd et al., 1999). The Voronoi cells are further assumed to be the OD regions observed in the MIP map. It is also possible to compute a Voronoi diagram at a finest spatial resolution including crossings as relevant features for traffic analysis.

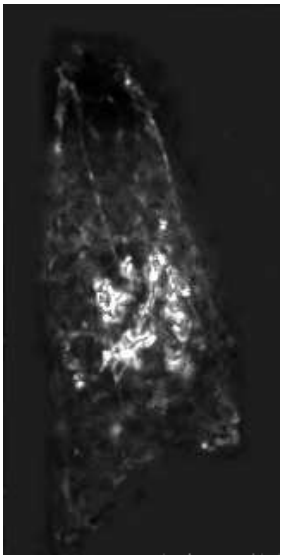


Figure 2: MIP map extracted from the image sequence shown in Fig. 1.

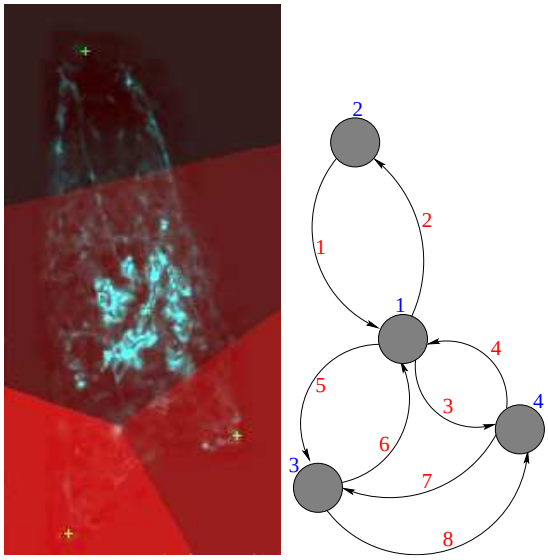


Figure 3: Left: partition of the compartment and surroundings observed in the image sequence shown in Fig. 1 by using a Voronoi decomposition. The different regions appear in red, their centers are labeled in green, and the MIP map is in the background; right: the corresponding graph; the vertex numbers are labeled in blue, while the edge numbers are labeled in red.

In order to partition the regions of interest within the cell, the expert can also manually define the centers of the Voronoi diagram if required. This diagram is then computed using the *qhull* library (Barber et al., 1996). A segmentation for the cell observed in the image sequence of Fig. 1 is typically illustrated in Fig. 3 where the centers appear in green and the different regions appear in red, while the MIP map is depicted in the background. In this figure, the centers were fixed to represent the Golgi Apparatus, and the three possible end-points of the cell.

The Voronoi diagram is also described by an adjacency graph (Fig. 3, right) and then consistent with the NT concept used for tracking. The different Voronoi cells represent the set of vertices V while the boundaries between the cells represent the set of edges E. We introduce two edges between two neighbouring cells in order to analyse trafficking in both directions.

Given the graph $\mathcal{G}$, the next step consists in extracting the data to apply the NT approach, i.e. estimating the number of vesicles that move from one Voronoi cell to another one during the whole image sequence.

2.2 Temporal Estimation of the Number of Moving Vesicles

We want to know exactly how many vesicles are moving from one Voronoi cell to another one at each time step. Our idea is to compute the difference of the number of vesicles observed at two consecutive time steps in each neighbouring region, and then to infer the exact number of vesicles that crosses each common boundary. Nevertheless, computing the difference of vesicles in each region involves image segmentation, a hard task since many similar objects overlap. By applying NT, we circumvent the problem since a crude partition of the image is only needed. In what follows, we assume that the level of fluorescence is proportional to the number of Rab6 proteins at each pixel. So the difference of image intensity at two time steps represents the difference of the number of Rab6 proteins in each region. In practice, the background corresponding to the Golgi apparatus and to the cytosol diffusion is first removed during a preprocessing step (Boulanger et al., 2006) for better performance. We illustrate this concept on a simple example explained below.

We consider the fluorescence exchanges at the vertex 1 in the graph shown in Fig. 3. Let $Z_{v,t}$ be the total amount of fluorescence in the complete Voronoi region corresponding to the vertex v at time t, and let $Y_{e,t}$ be the level of fluorescence to be determined on edge e at time t:

$$\begin{aligned} Z_{1,t+1} - Z_{1,t} &= Y_{1,t+1} - Y_{2,t+1} + Y_{4,t+1} - Y_{3,t+1} \\ &\quad + Y_{6,t+1} - Y_{5,t+1}. \end{aligned}$$

This equation can be extended to all vertices: let $\Delta\mathbf{Z}$ be the $n \times t$ matrix corresponding to the difference of

Table 1: Definition of the matrix **M** corresponding to the graph shown in Fig. 3.

	vertices			
edges	1	2	3	4
1	1	-1	0	0
2	-1	1	0	0
3	-1	0	0	1
4	1	0	0	-1
5	-1	0	1	0
6	1	0	-1	0

fluorescence in each region between two consecutive time steps, with n the number of regions and t the number of images in the sequence. Let $\mathbf{Y}$ be the $r \times t$ matrix representing the level of fluorescence that fluctuates from one region to another at each time, with r denoting the number of edges. We define $\mathbf{M}$ as the so-called "neighbourhood $n \times r$ matrix" composed of ternary elements $m = \{-1, 0, 1\}$ that links the regions according to the neighbourhood relationships. For example, in Fig. 3, $\mathbf{M}$ is defined as shown in Tab. 1. Then, we have:

$$\Delta\mathbf{Z} = \mathbf{M}\mathbf{Y} \quad (1)$$

Our aim is to estimate $\mathbf{Y}$ with $r > n$ given $\Delta\mathbf{Z}$, so to solve an under-constrained problem. Additional constraints are necessary for solving (1). First, we assume that all the components of $\mathbf{Y}$ are positive since the edges are unidirectional. In addition, the $\Delta\mathbf{Z}$ rows are assumed to be i.i.d., and we naturally choose the $\mathcal{L}^2$ distance. Finally, we propose to solve the following optimization problem:

$$\widehat{\mathbf{Y}} = \min_{\mathbf{Y}} \| \Delta\mathbf{Z} - \mathbf{M}\mathbf{Y} \|^2 \text{ subject to} \mathbf{Y} \geq \mathbf{0}.$$

This optimization problem leads to an estimation of $\mathbf{Y}$. To improve the solution, we also introduce an additional constraint based on the idea of *parsimony* (see (Tibshirani, 1996; Candes and Tao, 2007)). Actually, each row of $\widehat{\mathbf{Y}}$ corresponds to fluorescence exchanges on edges during the whole image sequence. In what follows, we want to check if the estimation of $\mathbf{Y}$ is improved when the traffic on some edges is removed, especially on edges for which a very low traffic is observed. Accordingly, $\mathbf{Y}$ is split into positive rows $\mathbf{Y}_l$ and rows with zero values $\mathbf{Y}_{n-l}$. The minimization can be then modified as follows:

$$(\hat{l}, \widehat{\mathbf{Y}}) = \min_{l, \mathbf{Y}_l} \| \Delta\mathbf{Z} - \mathbf{M}_l\mathbf{Y}_l \|^2 + \rho l, \text{subject to} \mathbf{Y} \geq \mathbf{0},$$

where the second term encourages the selection of few edges with l denoting the number of non-zero rows in $\mathbf{Y}$, ρ a balance term, $\mathbf{Y}_l$ the $(r-l) \times t$ matrix corresponding to $\mathbf{Y}$ restricted to rows with significant measurements (non zero), $\mathbf{M}_l$ the neighbourhood matrix that matches $\mathbf{Y}_l$, and $\Delta\mathbf{Z}$ denoting the difference of fluorescence in each region between two consecutive time steps.

In practice, we propose the following *greedy algorithm* for minimization:

1. compute $\widehat{\mathbf{Y}}_l = \min_{\mathbf{Y}_l \geq \mathbf{0}} \| \Delta\mathbf{Z} - \mathbf{M}_l\mathbf{Y}_l \|^2$,
2. compute $e = \| \Delta\mathbf{Z} - \mathbf{M}_l\widehat{\mathbf{Y}}_l \|^2 + \rho l$,
3. remove the row l' in $\widehat{\mathbf{Y}}_l$ that contains the higher number of 0 values,
4. update the matrices $\mathbf{Y}'$ and $\mathbf{M}'$ with $(l-1)$ components,
5. compute $\widehat{\mathbf{Y}}' = \min_{\mathbf{Y}' \geq \mathbf{0}} \| \Delta\mathbf{Z} - \mathbf{M}'\mathbf{Y}' \|^2$,
6. compute $e' = \| \Delta\mathbf{Z} - \mathbf{M}'\widehat{\mathbf{Y}}' \|^2 + \rho(l-1)$,
7. accept $\widehat{\mathbf{Y}} = \widehat{\mathbf{Y}}'$ if $e' < e$,
8. if all rows were considered, stop the procedure, else go back to step 2.

Finally, depending on the microtubule network topology and the related Voronoi diagram, the expert can also forbid the fluorescence transfer between several regions if required. This option is explained in the next section.

2.3 Traffic Partially Known

Biological motivations, confirmed by the MIP map, can be exploited to prevent the displacements of vesicles from one region to another one. This can be performed by modifying the neighbourhood matrix $\mathbf{M}$. For instance, in Fig. 3, we assume that the expert knows that no vesicle is moving between region 1 and region 2. The matrix is therefore modified accordingly as

$$\mathbf{M}(:,1) = 0, \text{ and}$$
$$\mathbf{M}(:,2) = 0,$$

with $\mathbf{M}(:,1) = \{\mathbf{M}(1,1); \mathbf{M}(2,1); \ldots; \mathbf{M}(n,1)\}$. For the expert, this flexibility can be appropriate for real applications where interactions with the image is required, as demonstrated in our experiments.

2.4 Experiments

In this section, we propose a first set of experiments to evaluate the performance of the estimation procedure, to be exploited in the NT approach described in Section 3. In this experiment, the vesicles at the origin and destination regions are stocked, to take into account the difference of fluorescence.

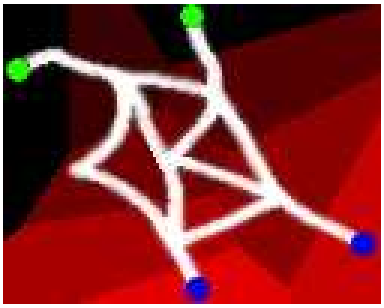


Figure 4: Network used for the first simulation. The microtubule network appears in white, the Voronoi cells are in red, the origin regions are labeled in green, and the destination regions are labeled in blue.

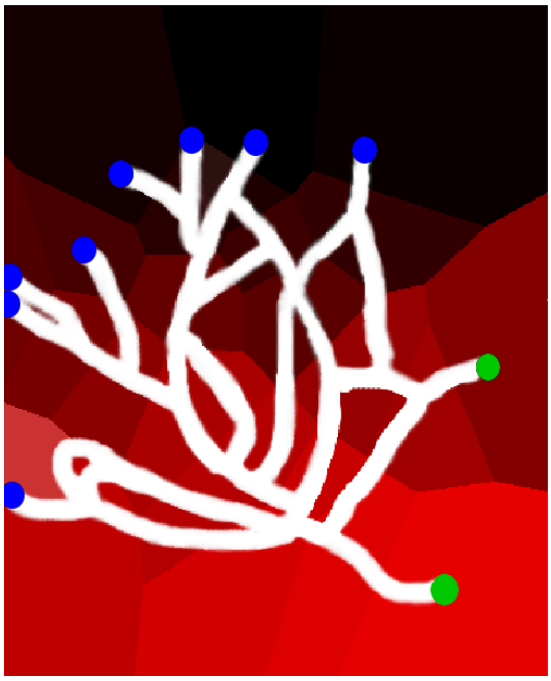


Figure 5: Network used for the second simulation. The microtubule network appears in white, the Voronoi cells are in red, the origin regions are labeled in green, and the destination regions are labeled in blue.

Table 2: Evaluation of the estimation of the traffic from the simulated network shown in Fig. 4.

Temporal tolerances	0	1	2	3
without *greedy algorithm*				
PFA	36%	9%	5%	4%
PFN	36%	9%	5%	4%
PGD	64%	91%	95%	96%
with *greedy algorithm*				
PFA	34%	6%	2%	1%
PFN	34%	6%	2%	2%
PGD	65%	94%	98%	99%

Table 3: Evaluation of the estimation of the traffic from the simulated network shown in Fig. 5.

Temporal tolerances	0	1	2	3
without *greedy algorithm*				
PFA	39%	9%	5%	5%
PFN	39%	11%	7%	7%
PGD	61%	89%	93%	93%
with *greedy algorithm*				
PFA	35%	5%	1%	1%
PFN	37%	8%	4%	4%
PGD	63%	92%	96%	96%

Two sequences are simulated based on the networks shown in Figs. 4 and 5, where the network appears in white, the Voronoi cells in red, the origin regions in green and the destination regions in blue. The simulations correspond to sequences of 1000 images, showing 2000 moving vesicles.
The performance of our estimation procedure described in Section 2.2 are measured using three criteria:

$$\text{PFA} = \frac{\text{number of false detections}}{\text{total real number of detections}},$$

$$\text{PFN} = \frac{\text{number of true detections not effected}}{\text{total real number of detections}},$$

$$\text{PGD} = \frac{\text{number of good detections}}{\text{total real number of detections}},$$

where PFA denotes the Probability of False Alarms, PFN the Probability of False Negatives, and PGD the Probability of Good Detections.

A slight temporal shifting between the estimation results and the “ground truth” is observed. That is why the results are presented with different temporal tolerances. For instance, a temporal tolerance equal to δt means that the estimation results are compared with a shifting in $[-\delta t, \ldots, \delta t]$ to the ground truth. The temporal estimations of the number of moving vesicles in the simulations based on the networks shown in Figs. 4 and 5 are given in Tabs. 2 and 3. In these tables, the results obtainedare shown with and without using the *greedy algorithm*.

Clearly, with a slight temporal tolerance, the estimated results are very close to the ”ground truth”. In addition, it worth noting that we only use temporal averages for NT, so the shifting will not be crucial for OD pairs estimation. Moreover, it is also confirmed that the *greedy algorithm* significantly improves the estimation results.

3 NETWORK TOMOGRAPHY

As explained in Section 2.1 and illustrated in Fig. 3, a region within the cell (e.g. compartment) can be represented by a graph corresponding to a Voronoi diagram, where the centers of the Voronoi cells correspond to regions of interest. The graph $\mathcal{G}(E,V)$ is defined by n vertices and r edges, where E denotes the set of edges, and V the set of vertices. A connection between two vertices is also called a path, and each path consists of one or more edges. In the NT-based approach, the data is the number of objects detected as going from one vertex to another vertex in the graph. Based on these measurements, the new goal is to estimate how many vesicles coming from an origin vertex go to a destination vertex along a path, in the set of all possible OD pairs in the graph, that is $c = n(n-1)$ OD pairs. This problem is then similar to determine the source-destination traffick based on link measurements in computer networks (Vardi, 1996). In this approach, it is not necessary to track an object through a dynamic scene, but just to determine when an object reaches a vertex, which is generally easier than estimating a continuous trajectory.

3.1 Problem Solving

More formally, let $X_{j,t}$, $j = 1,\cdots,c$, be the quantity of "transmitted" fluorescence on the OD pair j at time t. The measurements $\mathbf{Y}_t = (Y_{1,t},\ldots,Y_{r,t})^T$ are computed as explained in Section 2. The inherent randomness of the measurements motivates the adoption of a statistical approach. Now, we reasonably assume that the whole traffic is temporally distributed as a Poisson process, $X_{j,t} \sim \text{Poisson}(\lambda_j)$. In this traffic flow problem, we then assume the following model:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t, \tag{2}$$

where $\mathbf{X}_t = (X_{1,t},\ldots,X_{c,t})^T$, and $\mathbf{A}$ denotes a $r \times c$ routing matrix which binary elements $A_{ij} = 1$ if edge i is in the path for the OD pair j, and 0 otherwise.

For illustration, if we consider the simple example shown in Fig. 3, some rows of the matrix $\mathbf{A}$ are presented in Tab. 4. Typically, the number c is greater than r, and the problem is then under-constrained. Additional constraints are necessary for solving this inverse problem. First, (Vardi, 1996) proposed to introduce constraints related to the assumption that the traffic is temporally Poisson distributed. The NT method amounts then to estimating the values λ_j given the additional set of equations corresponding to temporal averages:

Table 4: Part of the matrix $\mathbf{A}$ corresponding to the graph shown in Fig. 3.

	edges					
OD pairs	1	2	3	4	5	6
$1 \rightarrow 2$	0	1	0	0	0	0
$1 \rightarrow 3$	0	0	0	0	1	0
$1 \rightarrow 4$	0	0	1	0	0	0
$2 \rightarrow 1$	1	0	0	0	0	0
$2 \rightarrow 3$	1	0	0	0	1	0
$2 \rightarrow 4$	1	0	1	0	0	0
...	...					

$$\begin{cases} \overline{Y_i} &= \sum_{k=1}^{c} A_{i,k}\lambda_k, & i = 1,\ldots,r, \\ cov(Y_i,Y_{i'}) &= \sum_{k=1}^{c} A_{i,k}A_{i',k}\lambda_k, & 1 \le i \le i' \le r. \end{cases}$$

This set of equations gives a system of $r(r+3)/2$ linear equations that forms an over-constrained problem that can be better solved with the conditions $\lambda_i \geq 0$. Moreover, in this application, the aim is not to obtain the number of vesicles that utilize each path, but to estimate the proportions of vesicles on each path. Hence, unlike previous methods (Vardi, 1996; Santini, 2000; Boyd et al., 1999), we impose the condition $\sum_{i=1}^{c}\lambda_i = 1$ as an additional constraint. The previous system can be written more compactly as:

$$\begin{pmatrix} \overline{\mathbf{Y}} \\ \mathbf{S} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}\Lambda, \tag{3}$$

where $\Lambda = (\lambda_1,\ldots,\lambda_c)^T$ contains the temporal mean of the traffic flow, $\mathbf{S} = \{cov(Y_i,Y_{i'})\}$ is the sample covariance matrix rewritten as a vector of length $r(r+1)/2$, and $\mathbf{B}$ is an $(r(r+1)/2) \times c$ matrix with the (i,i')th row of $\mathbf{B}$ being the element-wise product of row i and row i' of the matrix $\mathbf{A}$.

The system can be solved using the estimation-maximization (EM) method (Vardi, 1996; Santini, 2000) or the convex-projection algorithms (Boyd et al., 1999). In our case, we adapt a non negative mean square estimation which also provides a simple and reliable way to estimate the OD traffic $\widehat{\Lambda}$. For the implementation, our method is based on the *lsqnonlin* function from the Matlab Optimization toolbox. Note that a review of existing methods is also proposed in (Medina et al., 2002).

3.2 Origin-destination Regions Partially Known

When the expert specifies the origin or destination regions, the problem is better constrained and the solu-

tion is expected to be more relevant.

Typically, if we assume that the origins or destinations for the regions are known, this can be casted into additional hard constraints. If the Voronoi cell r is the single origin region, then all the OD pairs that have another Voronoi cell than r as origin have no longer meaning. So all that OD pairs can be ignored. Hence, let $\mathcal{R}$ be all the OD pairs that have r for origin. Then, if $\mathcal{O}$ denotes the set of all OD pairs, $\mathbf{A}$ can be modified as

$$\mathbf{A}(:, \mathcal{O} \setminus \mathcal{A}) = 0,$$

with $\mathbf{A}(:, \mathcal{O} \setminus \mathcal{A}) = \{\mathbf{A}(1, \mathcal{O} \setminus \mathcal{A}); \mathbf{A}(2, \mathcal{O} \setminus \mathcal{A}); \ldots; \mathbf{A}(r, \mathcal{O} \setminus \mathcal{A})\}$. The same modeling can be applied for imposing additional origin or destination regions.

4 EXPERIMENTAL RESULTS

In this section, we propose three experiments to demonstrate the ability and the limits of the NT-based approach applied to a real image sequence. All these experiments are tested by considering the sequence shown in Fig. 1. This sequence is composed of 900 images coming from a fast 4D deconvolution microscopy (wide-field) process (Sibarita et al., 2006). In this sequence, the background was removed during a preprocessing step. The estimated results are reported in Figs. 6 and 7. In these figures, the Voronoi cells are represented in red, while the MIP map is shown in the background by transparency. The different estimated OD pairs appear as colored arrows, and the corresponding colored numbers at the right top of the figures are the estimated proportions of moving vesicles for each OD pair.

A first experience was carried out with a crude segmentation, without imposing origin or destination regions. The results are shown in Fig. 6 (left). According to the expert-biologists, the vesicles are moving from the Golgi Apparatus (the central region) to end-points located at the periphery of the cell (corresponding to the three other regions). But, in this experience, the traffic is estimated going from end-points to end-points, which is not consistent with prior knowledge. That is why we impose, in a second experiment, (Fig. 6, right image), the central region to be the origin Voronoi cell. The results obtained with this additional constraint correspond to trafficking from the Golgi Apparatus to the end-points. In that case, the traffic tends to be quite uniform for all the end-points.

In another experiment corresponding to another partition of the image shown in Fig. 7 (left), the previous central Voronoi cell is divided into several cells,

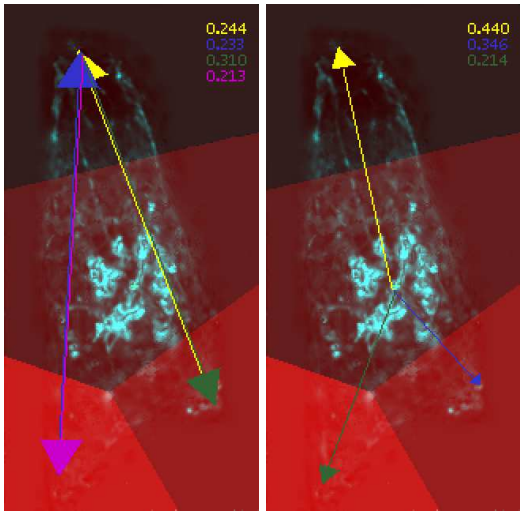


Figure 6: Results obtained by applying the NT-based approach on the sequence of the Fig. 1. The arrows represent the estimated OD pairs, and the corresponding colored numbers at the top right represent traffic proportions. Left: no origin region is imposed; right: the central region is imposed to be an origin region.

and they are all constrained to be origin regions. The estimated traffic from these origin regions to the end-points corresponds to proportions similar to proportions estimated in the previous experiment. In addition, the estimated traffic seems to be isotropic, i.e. there is no particular directions for traffic.

Finally, an experience is conducted with the same constrained origin region than the first experiment, but with one more end-point at the top of the image, and with intermediate Voronoi cells between the origin and the destinations (Fig. 7, right). Although the destination cells are not labeled, the whole trafficking is estimated from the Golgi Apparatus to the end-points. In addition, the sum of estimated proportions of the traffic towards the two regions at the top of the image is quite similar to the estimated proportions of the traffic towards the region at the top of the image in the first experiment. However, the estimated proportions of traffic towards the regions located at the bottom of the image are different from the estimated proportions of traffic towards the same regions in the first experience.

5 CONCLUSIONS

In this paper, we propose several contributions: *i*) definition of a graph by partitionning the image using a Voronoi diagram; *ii*) temporal estimation of moving vesicles; *iii*) application of the NT concept to real image sequences. The results obtained on the real image

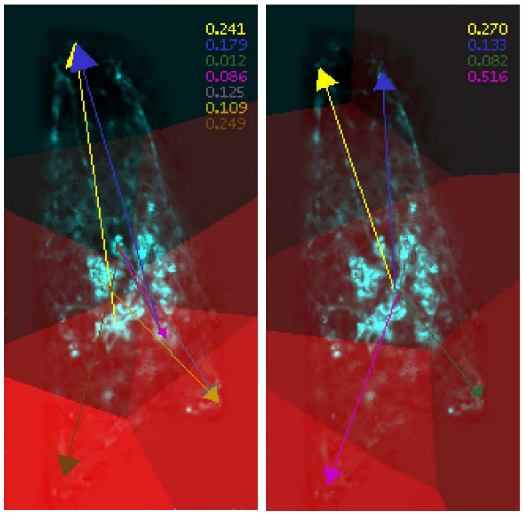


Figure 7: Results obtained by applying the NT-based approach on the sequence of Fig. 1. The arrows represent the estimated OD pairs, and the corresponding colored numbers at the top right represent traffic proportions. Left: all the central regions are imposed to be origin regions; right: the central region is imposed to be an origin region.

sequence suits the biological knowledge about the OD regions for the Rab6 trafficking. In our experiments, the proportions of vesicles for the OD pairs given by the NT procedure represent new tools for biologists. It can be applied to understand other trafficking problems where many objects are moving. Actually, the main limit is related to image partition yet, which can be arbitrary. Indeed, although the expert defines the centers of Voronoi cells with biological knowledge, the segmentation remains very crude for representing the regions of interest. Actually, the MIP map is the only tool available to define these regions, but is not enough accurate. For future work, it will be necessary to apply the NT-based approach on more relevant regions. A possible way is to extract the microtubule network and consider it as a graph for applying the NT procedure. Moreover, it is established that the fluorescence decreases with time, which is neglected in our modeling since we exploit the difference of fluorescence between two time steps. However, it is well known that the vesicles diffuse also in the cytosol. This could be considered in future work by introducing this phenomenon in the estimation process of the data to improve the results.

REFERENCES

Anderson, C., Georgiou, G., Morrison, I., Stevenson, G., and Cherry, R. (1992). Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charged-coupled device camera. low-density lipoprotein and influenza virus receptor mobility at 4 degrees c. *Journal of Cell Science*, 101:415–425.

Barber, C., Dobkin, D., and Huhdanpaa, H. T. (1996). The quickhull algorithm for convex hulls. *ACM transactions on mathematical software*, 22(4):469–483.

Bechar, I. and Trubuil, A. (2006). A bidimensional signal processing approach to vesicle trafficking analysis in 3d+t fluorescence videomicroscopy. *Proc. of the MICCAI Workshop on Microscopic Image Analysis and Application to Biology*.

Boulanger, J., Kervrann, C., and Bouthemy, P. (2006). Estimation of dynamic background for fluorescence video-microscopy. In *Proc. of ICIP'2006*, Atlanta.

Boyd, J. E., Meloche, J., and Vardi, Y. (1999). Statistical tracking in video traffic surveillance. In *Proc. of ICCV'99*, volume 1, pages 163–168.

Candes, E. and Tao, T. (2007). The dantzig selector: statistical estimation when p is much larger than n. *Annals of statistics*, to appear.

Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S., and Diot, C. (2002). Traffic matrix estimation: existing techniques and new directions. In *SIGCOMM '02: Proc. Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pages 161–174, New York, USA.

Pécot, T., Boulanger, J., Kervrann, C., and Bouthemy, P. (2007). Network tomography for trafficking simulation and analysis in fluorescence microscopy imaging. In *Proc. of IEEE ISBI'2007*, pages 268–271, Arlington.

Racine, V., Hertzog, A., Jouaneau, J., Salamero, J., Kervrann, C., and Sibarita, J. (2006). Mutliple target tracking of 3d fluorescent objects based on simulated annealing. In *Proc. of IEEE ISBI'2006*.

Santini, S. (2000). Analysis of traffic flow in urban areas using web cameras. In *Fifth IEEE Workshop on Applications of Computer Vision*, volume 596, pages 140–145.

Sbalzarini, I. and Koumoutsakos, P. (2005). Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*, 151:182–195.

Sibarita, J., Racine, V., and Salamero, J. (2006). Quantification of membrane trafficking on a 3d cytoskeleton network in living cells. In *Proc. of IEEE ISBI'2006*.

Smal, I., Niessen, W., and Meijering, E. (2007). Advanced particle filtering for multiple object tracking in dynamic fluorescence microscopy images. In *Proc. of IEEE ISBI'2007*, pages 1048–1051, Arlington.

Thomann, D., Dorn, J., Sorger, P., and Danuser, G. (2003). Automatic fluorescent tag localization ii: improvement in super-resolution by relative tracking. *Journal of Microscopy*, 211(3):230–248.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288.

Vardi, Y. (1996). Network tomography: Estimation of source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91(433):365–377.

BIOSIGNAL-BASED COMPUTING BY AHL INDUCED SYNTHETIC GENE REGULATORY NETWORKS

From an in vivo Flip-Flop Implementation to Programmable Computing Agents

T. Hinze, T. Lenser, N. Matsumaru, P. Dittrich
Bio Systems Analysis Group, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 1–4, D-07743 Jena, Germany
{*hinze,thlenser,naoki,dittrich*}*@minet.uni-jena.de*

S. Hayat
Computational Biology Group, Universität des Saarlandes, Im Stadtwald, PF 15 11 50, D-66041 Saarbrücken, Germany
s.hayat@bioinformatik.uni-saarland.de

Keywords: Biologically inspired computing, gene regulatory networks, bistable toggle switch, NP complete problem.

Abstract: Gene regulatory networks (GRNs) form naturally predefined and optimised computational units envisioned to act as biohardware able to solve hard computational problems efficiently. This interplay of GRNs via signalling pathways allows the consideration as well as implementation of interconnection-free and fault tolerant programmable computing agents. It has been quantitatively shown in an *in vivo* study that a reporter gene encoding the green fluorescent protein (*gfp*) can be switched between high and low expression states, thus mimicking a NAND gate and a RS flip-flop. This was accomplished by incorporating the N-acyl homoserine lactone (AHL) sensing lux operon from *Vibrio fischeri* along with a toggle switch in *Escherichia coli*. *gfp* expression was quantified using flow cytometry. The computational capacity of this approach is extendable by coupling several logic gates and flip-flops. We demonstrate its feasibility by designing a finite automaton capable of solving a knapsack problem instance.

1 INTRODUCTION

Molecular biosignals of GRNs are suited for computing purposes. Beyond error-prone molecular computers *in vitro*, the idea of computing *in vivo* becomes more and more attractive (Kobayashi et al., 2004; Weiss et al., 1999). Motivated by the presence of naturally predefined functional units found in microorganisms, their exploration towards computing agents seems promising. GRNs can be viewed as computational devices of pro- and eukaryotic cells, triggering and directing responses to external inputs influenced by genetically stored information. Emerged from an evolutionary optimisation, they form reliable modularised systems with obvious advantages in carrying out massively parallel calculations.

Recently, more than $6,000$ transcription factors acting as signal inducers or transmitters in GRNs have been listed in the public Transpath and Transfac database. Selection of GRN candidates in concert with their ability for composition via specific receptor-controlled interfaces is of particular interest for creating computing agents. Following the pioneering implementation of a bistable toggle switch *in vivo* (Gardner et al., 2000), we could confirm its practicability in a previous study (Hayat et al., 2006). Two extensions in the experimental setup were investigated: Firstly, the effects of isopropyl-β-D-thiogalactopyranoside (IPTG) and AHL as appropriate intercellular inducers for flip-flop setting were shown. Secondly, flow cytometry was used to quantitatively measure protein concentrations within the flip-flop implementation. Section 3 briefly explains underlying methods and materials, while section 4 describes the experiment. Based on these experimental results, we identify GRNs able to function as NAND gates and RS flip-flops (section 5). For these computational units, we present a homogeneous dynamical model using Hill kinetics (Mestl et al., 1995). Unit composition is exemplified by design of a finite automaton for a knapsack problem instance (section 6). We transfer this automaton into a minimised logic circuit consisting of interacting NAND gates. A subsequent simulation of the corresponding computing agent on the GRN level demonstrates a constructive approach towards programmable *in vivo* computers encouraged by the Turing completeness of (bio)chemical reaction networks (Magnasco, 1997).

2 RELATED WORK

First ideas about principles of interconnection-free biomolecular computation were introduced in (Aoki et al., 1992; Thomas, 1991). Along with the growing knowledge in genetics and proteomics, the imagination of computing *in vivo* came into the focus of scientific research (Weiss et al., 1999). Several approaches address engineering of GRNs using predefined network motifs (Guido et al., 2006; Kaern et al., 2003; Kobayashi et al., 2004). Inspired by advantageous properties of specific GRNs for computational tasks, construction and implementation of genetic circuits has been successfully explored (Gardner et al., 2000; Sprinzak and Elowitz, 2005; Yokobayashi et al., 2004). Recently, these circuits have been used to form logic gates, clocks (oscillators), switches, or sensors. As a next step, their combination to achieve higher integrated problem-specific designs will be investigated. A DNA-based solution to the knapsack problem *in vitro* was introduced in (Hinze et al., 2002). An alternative approach *in vivo* generates a variety of encoding plasmids that are translated into proteins (Henkel et al., 2007).

3 BIOLOGICAL PRINCIPLES AND PREREQUISITES

3.1 Quorum Sensing and Autoinduction via AHL

In quorum sensing, bacterial species regulate gene expression based on cell-population density (Miller and Bassler, 2001). An alteration in gene expression occurs when an intercellular signalling molecule termed autoinducer, produced and released by the bacterial cells reaches a critical concentration. Termed as quorum sensing or autoinduction, this fluctuation in autoinducer concentration is a function of bacterial cell-population density. *Vibrio fischeri*, a well studied bacterium, colonises the light organs of a variety of marine fishes and squids, where it occurs at very high densities ($10^{10}\frac{\text{cells}}{\text{ml}}$) and produces light. The two genes essential for cell density regulation of luminescence are: luxI, which codes for an autoinducer synthase (Schaefer et al., 1996); and luxR, which codes for an autoinducer-dependent activator of the luminescence genes. The luxR and luxI genes are adjacent and divergently transcribed, and luxI is the first of seven genes in the luminescence or lux operon. LuxI-type proteins direct AHL synthesis while LuxR-type proteins function as transcriptional regulators that are capable of binding AHL signal molecules. Once formed, LuxR-AHL complexes bind to target promoters of quorum-regulated genes. Quorum sensing is now known to be widespread among both Gram-positive and Gram-negative bacteria.

3.2 Bioluminescence in *Vibrio fischeri*

Bioluminescence in general is defined as an enzyme catalysed chemical reaction in which the energy released is used to produce an intermediate or product in an electronically excited state, which then emits a photon. It differs from fluorescence or phosphorescence as it is not depended on light absorbed. The mechanism for gene expression and the structure of the polycistronic message of the lux structural genes in *Vibrio fischeri* have been thoroughly characterised (Hastings and Nealson, 1977). Briefly, there are two substrates, luciferin, which is a reduced flavin mononucleotide ($FMNH_2$), and a long chain ($7-16$ carbons) fatty aldehyde (RCHO). An external reductant acts via flavin mono-oxygenase oxidoreductase to catalyse the reduction of FMN to $FMNH_2$, which binds to the enzyme and reacts with O_2 to form a 4a-peroxy-flavin intermediate. This complex oxidises the aldehyde to form the corresponding acid (RCOOH) and a highly stable luciferase-hydroxyflavin intermediate in its excited state, which decays slowly to its ground state emitting blue-green light $h\nu$ with a maximum intensity at about 490nm:

$$\begin{aligned} &FMNH_2 + RCHO + O_2 \xrightarrow{\text{lucif.}} \\ &FMN + H_2O + RCOOH + h\nu \end{aligned} \quad (1)$$

3.3 Transcription Control by LacR and λCI Repressor Proteins

Escherichia coli cells repress the expression of the lac operon when glucose is abundant in the growth medium. Only when the glucose level is low and the lactose level is high, the operon is fully expressed. The Lac repressor LacR is a 360 residue long protein that associates into a homotetramer. It contains a helix-turn-helix (HTH) motif through which it interacts with DNA. This interaction represses transcription by hindering association with RNA polymerase and represents an example of “combinatorial control” widely seen in prokaryotes and eukaryotes (Buchler et al., 2003). The CI repressor of bacteriophage lambda is the key regulator in lambda’s genetic switch, a bistable switch that underlies the phage’s ability to efficiently use its two modes of development (Ptashne, 1992).

3.4 Flow Cytometry

Flow cytometry refers to the technique where microscopic particles are counted and examined as they pass in a hydro-dynamically focused fluid stream through a measuring point surrounded by an array of detectors. Previously, flow cytometry analyses were performed by us using a BD LSRII flow cytometer equipped with 405nm, 488nm and 633nm lasers. 488nm laser was used for *gfp* and yellow fluorescent protein (*yfp*) quantification.

4 IMPLEMENTATION OF A COMPUTATIONAL UNIT *IN VIVO*

4.1 Experimental Setup

We have shown that an *in vivo* system (Hayat et al., 2006) can potentially be used to mimic a RS flip-flop (Huffman, 1954; Morris Mano, 1991) and have quantified its performance using flow cytometry. The presence or absence of the inducers IPTG or AHL in combination with temperature shift acts as an input signal. The toggle switch comprising of structural genes for reporter/output proteins fused to promoter regions that are regulated by input signals is visualised as a RS flip-flop, see figure 1. The functional modularity of the input and output circuits is maintained so that the artificial GRN used can be easily extended for future studies.

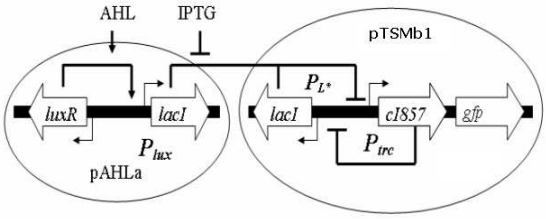


Figure 1: A schematic diagram of an AHL biosensor module interfaced with the genetic toggle switch adapted from (Hayat et al., 2006). The transgenic artificial GRN consists of a bistable genetic toggle switch (Gardner et al., 2000) which is interfaced with genes from the lux operon (Engebrecht and Silverman, 1984) of the quorum sensing signalling pathway of *Vibrio fischeri* (Schaefer et al., 1996).

This design endows cells with two distinct phenotypic states: where the λCI activity is high and the expression of lacI is low (referred to as high or 1 state), or where the activity of LacR is high and the expression of λCI is low (referred to as low or 0 state). *gfp* is expressed only in the high λCI/low LacR state. Figure 2 shows the experimental results obtained by flow cytometry.

4.2 Results and Discussion

For co-relational purposes, all experiments were conducted with both BL21 and Top10 strains of *Escherichia coli*. The concentration of IPTG used in all the experiments was 2mM and that of AHL was 1μM. Experiments conducted without the use of inducers, lead to an unreliable shifting of the states, signifying the use if inducers in a tightly, mutually regulated circuit. Further experiments conducted to understand the switching dynamics of the circuit revealed that in the current scenario, it was easier to switch from a high to a low state than vice versa. This discrepancy in switching behaviour is attributed to the differing modes of elimination of LacR and λCI repressor proteins. While switching from low to high state, the repression due to IPTG-bound Lac repressor needs to be overcome by cell growth. Switching from high to low state is effected by immediate thermal degradation of the temperature-sensitive λCI. Experiments were also conducted to test the sustainability of states. The plug and play property of the circuit was examined by employing *yfp* as the reporter gene instead of *gfp*. As shown in figure 2, the circuit could reliably mimic a RS flip-flop. The massive parallelism permissible by the use of large quantities of cells can compensate for the slow speed of switching. Further tests are to be performed to confirm this hypothesis.

5 DEFINITION OF COMPUTATIONAL UNITS

Artificial GRNs have been instrumental in elucidating basic principles that govern the dynamics and consequences of stochasticity in the gene expression of naturally occurring GRNs. The realisation as computational circuits infers inherent evolutionary fault tolerance and robustness to these modular units.

In a more abstract view, generalised building blocks adopted from previous *in vivo* studies can be identified. In terms of predefined GRN modules, they form computational units. We introduce two artificial GRNs for logic gates and describe their dynamical behaviour quantitatively by an ordinary differential equation model using Hill kinetics (Mestl et al., 1995). A variety of specific signalling proteins (transcription factors) given by their concentration over the time course enables communication between as well as coupling of computational units. Thus, circuit

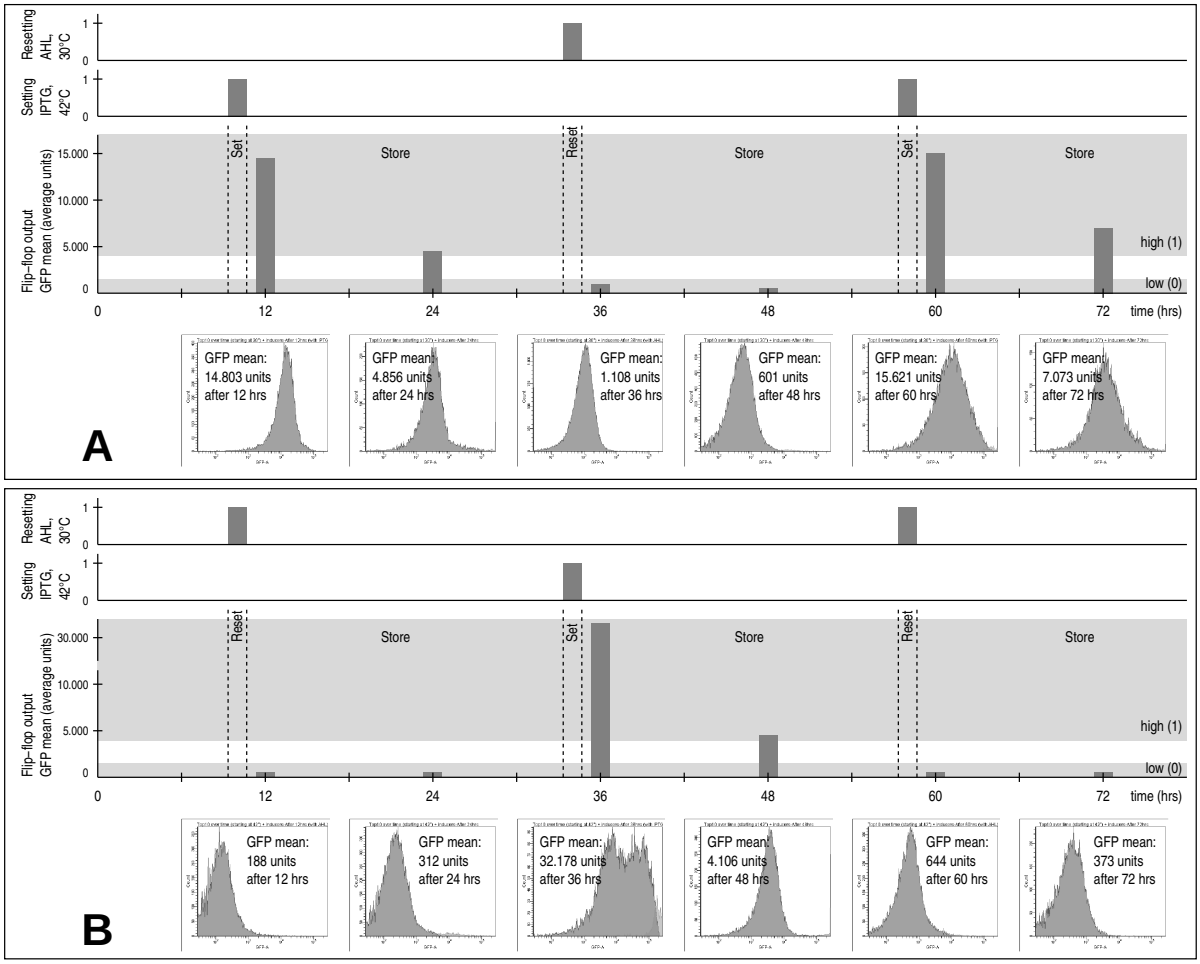


Figure 2: Inducer-dependent switching. Repeated activation and deactivation of the toggle switch based on inducers and temperature. Temperature was switched every 24 hours. Cells were incubated with inducers for 12 hours, followed by growth for 12 hours without inducers, initially kept at 30°C (**A**) and 42°C (**B**). The cells successfully switched states thrice.

engineering becomes feasible. Resulting computing agents can serve as templates for experimental setups. We distinguish computational units within a circuit by index i.

Hill kinetics represents a homogeneous approach to model cooperative and competitive aspects of interacting biochemical reaction networks. It formulates the relative intensity of gene regulations by sigmoid-shaped threshold functions h of degree m and threshold $\Theta > 0$ such that $x \geq 0$ specifies the concentration level of a transcription factor that activates resp. inhibits gene expression. Function value h then returns the normalised change in concentration level of the corresponding gene product. It reaches a concentration level of 50% iff $x = \Theta$.

Activation (upregulation) $\rightarrow$:

$$\mathrm{h}^+(x,\Theta,m) = \frac{x^m}{x^m + \Theta^m} \qquad (2)$$

Inhibition (downregulation) $\perp$:

$$\mathrm{h}^-(x,\Theta,m) = 1 - \mathrm{h}^+(x,\Theta,m) \qquad (3)$$

5.1 NAND Gate

Input: concentration levels of transcription factors x_i, y_i

Output: concentration level of gene product z_i

The dynamical system behaviour is defined by ordinary differential equations 4, 5, and 6.

$$\dot{a}_i = \mathrm{h}^+(x_i,\Theta_{i1},m_{i1}) - a_i \qquad (4)$$

$$\dot{b}_i = \mathrm{h}^+(y_i,\Theta_{i2},m_{i2}) - b_i \qquad (5)$$

$$\dot{z}_i = 1 - \mathrm{h}^+(a_i,\Theta_{i3},m_{i3}) \cdot \mathrm{h}^+(b_i,\Theta_{i4},m_{i4}) - z_i \qquad (6)$$

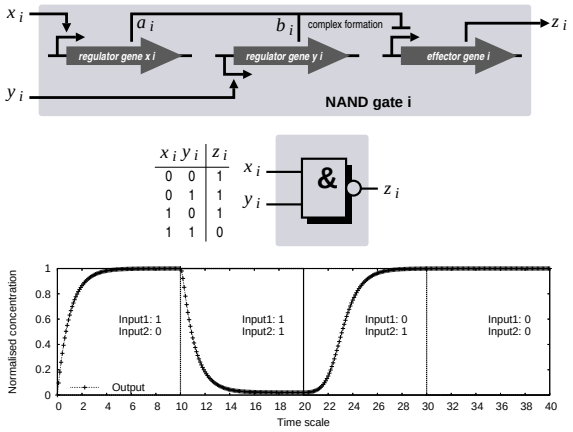


Figure 3: GRN mimicking a NAND gate. Dynamical behaviour shown for $m_{ij} = 2$, $\Theta_{ij} = 0.1$, $j = 1,\ldots,4$, $a(0) = 0$, $b(0) = 0$, $z(0) = 0$ (Copasi ODE solver simulation).

5.2 RS Flip-Flop

Input: concentration levels of transcription factors $\overline{S_i}, \overline{R_i}$

Output: concentration level of gene product Q_i

The dynamical system behaviour is defined by ordinary differential equations 7, 8, and 9.

$$\dot{a_i} = 1 - \mathrm{h}^+(b_i, \Theta_{i1}, m_{i1}) \cdot \mathrm{h}^-(\overline{S_i}, \Theta_{i2}, m_{i2}) - a_i \quad (7)$$

$$\dot{b_i} = 1 - \mathrm{h}^+(a_i, \Theta_{i3}, m_{i3}) \cdot \mathrm{h}^-(\overline{R_i}, \Theta_{i4}, m_{i4}) - b_i \quad (8)$$

$$\dot{Q_i} = \mathrm{h}^+(b_i, \Theta_{i1}, m_{i1}) \cdot \mathrm{h}^-(\overline{S_i}, \Theta_{i2}, m_{i2}) - Q_i \quad (9)$$

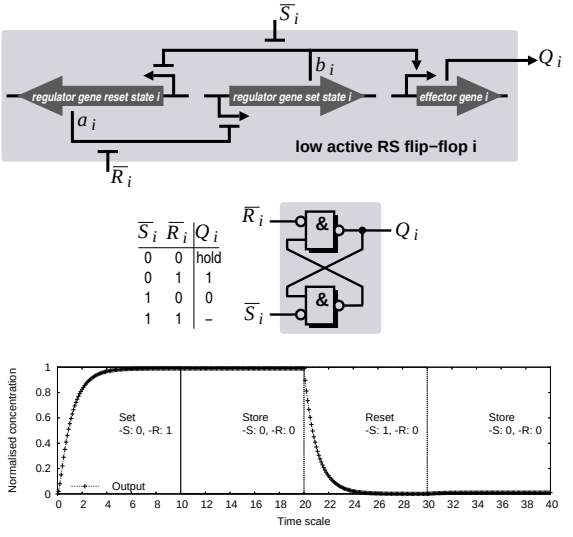


Figure 4: GRN mimicking a low active RS flip-flop. Dynamical behaviour shown for $m_{ij} = 2$, $\Theta_{ij} = 0.1$, $j = 1,\ldots,4$, $a(0) = 0$, $b(0) = 0$, $Q(0) = 0$ (Copasi ODE solver simulation).

A homologous model of a bistable toggle switch was introduced in (Gardner et al., 2000). In case of the forbidden input signalling $\overline{S_i} = 1$, $\overline{R_i} = 1$, the normalised concentrations of both proteins a_i and b_i converge to 0.5. By setting or resetting input signalling, the flip-flop restores.

6 AN ARTIFICIAL GRN FOR KNAPSACK PROBLEM SOLUTION

We demonstrate the feasibility of unit composition for solution of more complex problems, exemplified by the integer knapsack problem. It is known to be NP complete, defined by n natural numbers $a_1,\ldots,a_n$ representing weights of corresponding objects $1,\ldots,n$ and a reference weight b given by a natural number. Is there a subset $I \subseteq \{1,\ldots,n\}$ such that $\sum_{i\in I} a_i = b$? Brute force approaches enumerating the whole search space consider up to 2^n solution candidates.

Regarding a reaction network-based algorithm, the dynamic programming approach introduced in (Baum and Boneh, 1999) provides an appropriate framework to be adapted for our purpose. Here, the problem parameters are encoded into a directed graph $\mathcal{G} = (V,E)$ with a $(b+1) \times (n+1)$ grid of nodes $V = \{v_{(i,k)} \mid \forall i = 0,\ldots,b \,\forall k = 0,\ldots,n\}$ and edges $E \subset V \times V$ as follows: $E = \{(v_{(i,k)}, v_{(i,k+1)}) \mid \forall i = 0,\ldots,b \,\forall k = 0,\ldots,n-1\} \cup \{(v_{(i,k)}, v_{(i+a_i,k+1)}) \mid \forall i = 0,\ldots,b : i + a_i \leq b \,\forall k = 0,\ldots,n-1\}$. The answer to the knapsack problem is yes iff there exists a path through $\mathcal{G}$ from $v_{(0,0)}$ to $v_{(b,n)}$.

As an example for GRN network composition, we choose the problem instance $n = 3, a_1 = 3, a_2 = 1, a_3 = 2, b = 3$. Upper part of figure 5 illustrates graph $\mathcal{G}$ in this case. Having in mind the presence of a separate signal generator module, we can interpret $\mathcal{G}$ as a finite automaton. Assume that the generator module continually disseminates transmitter substances representing binary strings. Computing agents are able to receive these substances in parallel. So, one agent randomly obtains one binary string for subsequent processing in terms of input data. Acting as a (finite) automaton, the agent can reach a final state which is coupled to the expression of *gfp*.

In this example, the agent checks whether or not the binary string as a candidate solution meets the answer "yes". Therefore, the first three input bits are interpreted as presence (1) or absence (0) of an object corresponding to the bit position. Accordingly, binary strings 011 and 100 lead to the positive answer. In order to construct the finite automaton from grid $\mathcal{G}$, we

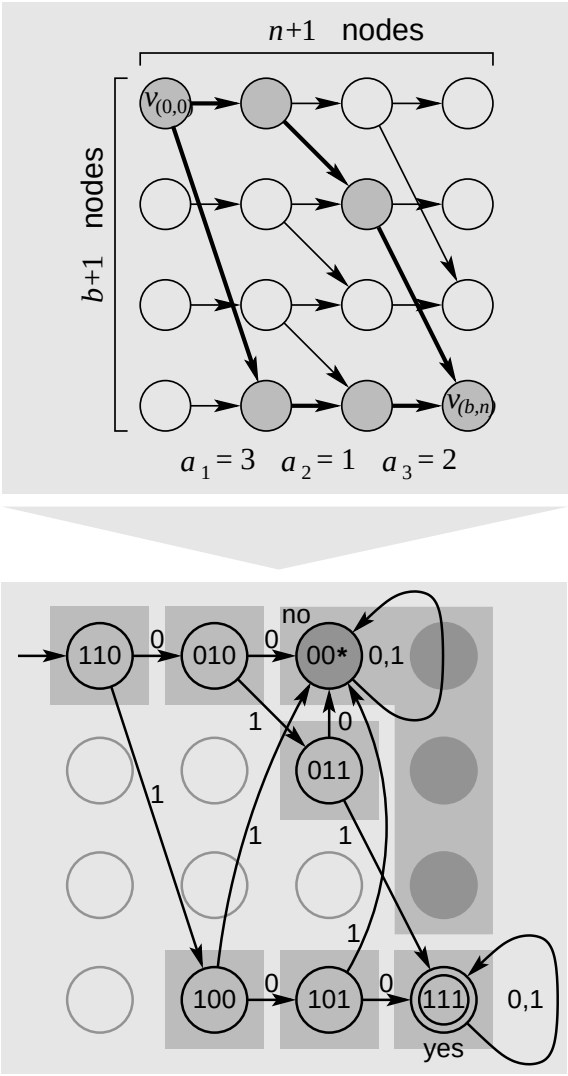


Figure 5: Construction of a deterministic finite automaton (lower part) from the dynamic programming approach (upper part).

transform the nodes into states whereas $v_{(0,0)}$ is the initial and $v_{(b,n)}$ the final state. Unreachable states are removed, and all four states that yield the answer "no" are subsumed into one, see lower part of figure 5. Finally, we obtain seven states. For preparation of the NAND-based network as template for GRN composition, we term each state by a three bit binary number $b_1b_2b_3$. With regard to a minimal number of NAND gates in the final circuit, the Gray code schema is used for state numbering. So, each transition flips at most one bit. Additionally, one state may carry a one-bit don't-care term marked by $*$ that arbitrarily embodies 1 or 0. Using Karnaugh optimisation, the binary number $b'_1b'_2b'_3$ for the new state based on the previous state $b_1b_2b_3$ and the input bit x is defined by:

$$
\begin{aligned}
b'_1 &= (b_1 \wedge b_2 \wedge b_3) \vee (b_1 \wedge \overline{b_2} \wedge \overline{x}) \\
&\quad \vee (b_2 \wedge b_3 \wedge x) \vee (b_1 \wedge b_2 \wedge x) \qquad (10)\\
b'_2 &= (b_1 \wedge b_2 \wedge b_3) \vee (b_1 \wedge b_2 \wedge \overline{x}) \\
&\quad \vee (b_1 \wedge b_3 \wedge \overline{x}) \vee (\overline{b_1} \wedge b_2 \wedge x) \qquad (11)\\
b'_3 &= b_3 \vee (\overline{b_2} \wedge \overline{x}) \vee (\overline{b_1} \wedge x) \qquad (12)
\end{aligned}
$$

The resulting NAND-based network shown in the upper part of figure 6 was directly derived from these transition rules. After backtransformation of this network into the description level of coupled GRNs, we obtain an artificial biosignalling system consisting of 115 interacting activation resp. repression pathways. The lower part of figure 6 depicts the normalised concentrations of the signalling substances encoding state bits b_1 and b_3 over the time course for three state transitions. Effects of signal weakening, also observed in laboratory studies *in vivo* (see figure 2), are reflected by the model based on Hill kinetics.

7 CONCLUSIONS

This paper addresses three strongly interconnected aspects of biomolecular computing based on biosignals processed by GRNs: wetlab implementation of computational units (NAND gate, low active RS flip-flop) *in vivo*, homogeneous dynamical modelling of these units (Hill kinetics), and their composition to computing agents able to solve real world problems, initially shown in a conceptual study *in silico*. We incorporated AHL as an additional intercellular transcription factor suitable to switch logic gates. Despite being rather slow, they convince by their reliability as a requirement for scalability. Flow cytometry provides a promising method to visualise amounts of output proteins resulting from computing processes *in vivo*. Further studies will consider parallel interactions of GRNs.

ACKNOWLEDGEMENTS

This work is part of the ESIGNET project (Evolving Cell Signalling Networks *in silico*), which has received research funding from the European Community's Sixth Framework Programme (project no. 12789). Further funding from the Federal Ministry of Education and Research (BMBF, grant 0312704A) is acknowledged. We are very grateful to J.J. Collins for providing us with the plasmids and their sequences; to W. Pompe, G. Rödel, K. Ostermann, and L. Brusch

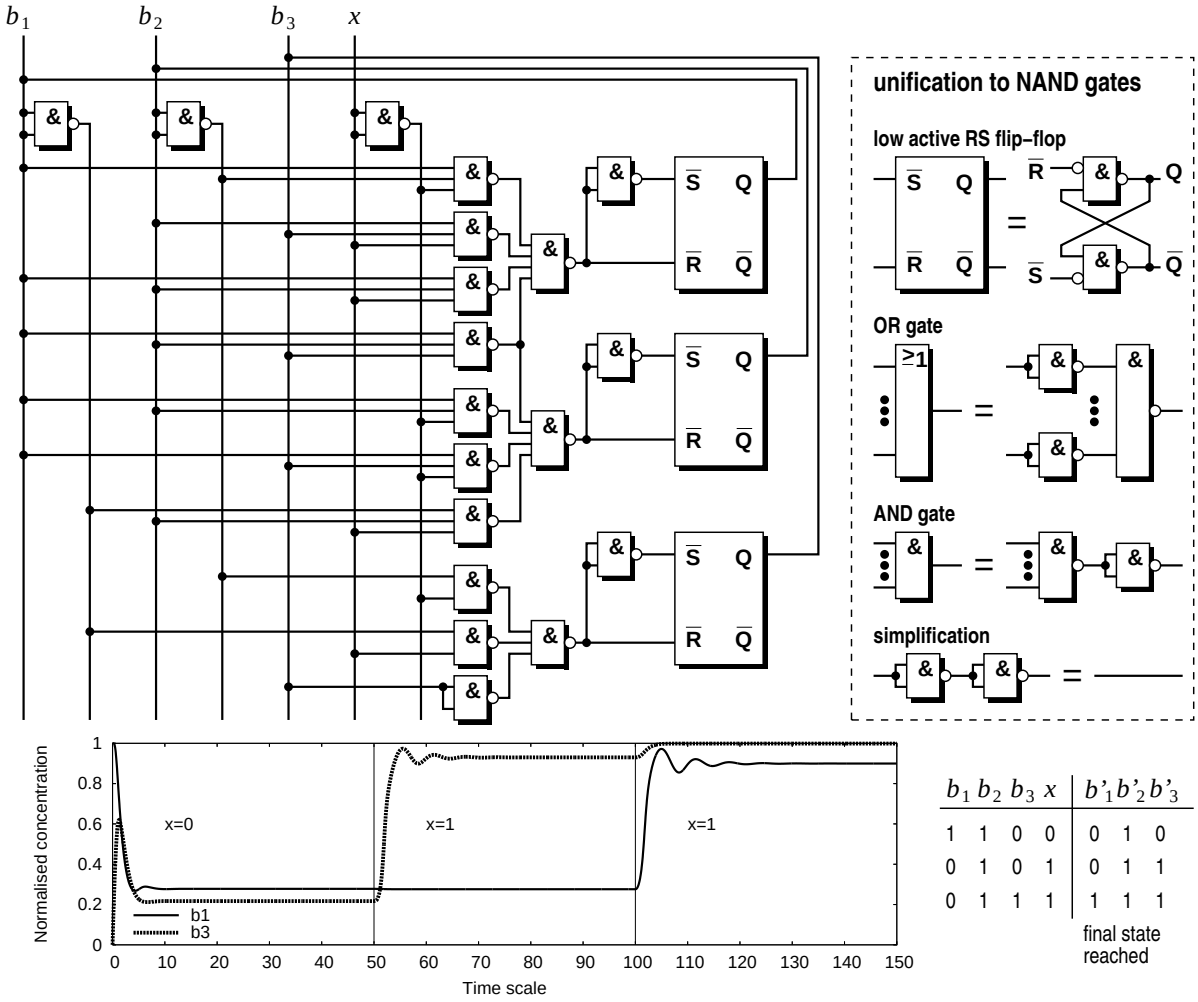


Figure 6: Minimal NAND-based logic circuit derived from the finite automaton and dynamical simulation of the system resulting from coupled GRNs (Copasi ODE solver). Up to 115 regulatory processes interact via signalling molecules (input/output substances of underlying 22 computational units). The diagram depicts the path $110 \xrightarrow{0} 010 \xrightarrow{1} 011 \xrightarrow{1} 111$ through the finite automaton states $b_1b_2b_3$. While state bit b_2 remains at the high level (not shown), b_1 and b_3 switch twice resp. once. Hill kinetic parameters were set as follows: $m_{ij} = 2$, $\Theta_{ij} \in [0.1, 0.3]$ in accordance with achieving stable system states corresponding to binary numbers. Normalised concentration levels between 0 and 0.3 are treated as binary 0, values between 0.7 and 1 represent binary 1.

from Dresden University of Technology for their scientific support and V. Helms from Saarland University for administrative support.

REFERENCES

Aoki, T., Kameyama, M., and Higuchi, T. (1992). Interconnection-free biomolecular computing. *IEEE Computer*, 25:41–50.

Baum, E. and Boneh, D. (1999). Running dynamic programming algorithms on a DNA computer. *DIMACS*, 44:77–86. Proceedings Second Meeting on DNA-Based Computers.

Buchler, N., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA*, 100(9):5136–5141.

Engebrecht, J. and Silverman, M. (1984). Identification of genes and gene products necessary for bacterial bioluminescence. *Proc. Natl. Acad. Sci. USA*, 81:4154–4158.

Gardner, T., Cantor, C., and Collins, J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature*, 403:339–342.

Guido, N., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C., Elston, T., and Collins, J. (2006). A bottom-up approach to gene regulation. *Nature*, 439:856–860.

Hastings, J. and Nealson, K. (1977). Bacterial bioluminescence. *Annu. Rev. Microbiol.*, 31:549–595.

Hayat, S., Ostermann, K., Brusch, L., Pompe, W., and Roedel, G. (2006). Towards in vivo computing: Quantitative analysis of an artificial gene regulatory network behaving as a RS flip-flop and simulating the system in silico. In *Proc. Bionetics Cavalese*. IEEE. in print.

Henkel, C., Baeck, T., Kok, J., Rozenberg, G., and Spaink, H. (2007). DNA computing of solutions to knapsack problems. *Biosystems*. in print.

Hinze, T., Hatnik, U., and Sturm, M. (2002). An object oriented simulation of real occurring molecular biological processes for DNA computing and its experimental verification. In *DNA Computing*, volume 2340 of *Lecture Notes in Computer Science*, pages 1–13.

Huffman, D. (1954). The synthesis of sequential switching circuits. *Journal of the Franklin Institute*, 257(3):161–190.

Kaern, M., Blake, W., and Collins, J. (2003). Engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, 5:179–206.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T., Cantor, C., and Collins, J. (2004). Programmable cells. *Proc. Natl. Acad. Sci. USA*, 101:8414–8419.

Magnasco, M. (1997). Chemical kinetics is Turing universal. *Physical Review Letters*, 78(6):1190–1193.

Mestl, T., Plahte, E., and Omholt, S. (1995). A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.*, 176:291–300.

Miller, M. and Bassler, B. (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.*, 55:165–199.

Morris Mano, M. (1991). *Digital Design*. Prentice-Hall, second edition.

Ptashne, M. (1992). *Genetic Switch: Phage Lambda and Higher Organisms*. Blackwell.

Schaefer, A., Val, D., Hanzelka, B., Cronan jr., J., and Greenberg, E. (1996). Generation of cell-to-cell signals in quorum sensing: Acyl homoserine lactone synthase activity of a purified vibrio fischeri LuxI protein. *Proc. Natl. Acad. Sci. USA*, 93:9505–9509.

Sprinzak, D. and Elowitz, M. (2005). Reconstruction of genetic circuits. *Nature*, 438:443–448.

Thomas, R. (1991). Regulatory networks seen as asynchronous automata: a logical description. *J. Theor. Biol.*, 53:1–23.

Weiss, R., Homsy, G., and Knight jr., T. (1999). Toward in-vivo digital circuits. *DIMACS*. Workshop on Evolution as Computation.

Yokobayashi, Y., Weiss, R., and Arnold, F. (2004). Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA*, 99:16587–16591.

AUTOMATIC SEGMENTATION OF CAPILLARY NON-PERFUSION IN RETINAL ANGIOGRAMS

Amit Agarwal, Jayanthi Sivaswamy
CVIT, International Institute of Information Technology, Hyderabad, India
amit@students.iiit.ac.in, jayanthi@iiit.net

Alka Rani
Aravind Eye Institute, Hyderabad, India
alka_bala@rediffmail.com

Taraprasad Das
LV Prasad Eye Institute, Hyderabad, India
tpd@lvpei.org

Keywords: Capillary Non-Perfusion, Retina, Extrema Pyramid, Disease detection.

Abstract: Capillary Non-Perfusion (CNP) is a condition in diabetic retinopathy where blood ceases to flow to certain parts of the retina, potentially leading to blindness. This paper presents a solution for automatically detecting and segmenting CNP regions from fundus fluorescein angiograms (FFAs). CNPs are modelled as valleys, and a novel multiresolution technique for trough-based valley detection is presented. The proposed algorithm has been tested on 40 images and validated against expert-marked ground truth. Obtained results are presented as a receiver operating characteristic (ROC) curve. The area under this curve is 0.842 and the distance of ROC from the ideal point $(0, 1)$ is 0.31.

1 INTRODUCTION

Diabetes is occurring in an ever increasing percentage in the world. Diabetes mellitus affects many organs of the body, and the eye is one of the organs that is affected relatively early (compared to the kidney). While diabetes affects all parts of the eye, the retina (retinopathy) is most commonly affected. Diabetic retinopathy progresses in phases. It starts with microaneurysms and superficial retinal hemorrhages (non-proliferative diabetic retinopathy; NPDR), progresses to accumulation of hard exudates in the posterior pole (diabetic maculopathy), and finally ends with new vessels in the surface of the retina and/or the optic disc (proliferative diabetic retinopathy; PDR). The underlying cause of the terminal event, the retinal new vessels, is retinal ischemia which manifests as areas of CNP that is most clearly seen in an FFA. These lesions appear as dark regions in the FFA images as shown in Fig. 1. If not treated in time, the CNP areas grow and spread across the entire retina. Large areas of non-perfusion lead to new vessel formation and bleeding into the vitreous cavity. These complications are responsible for severe visual loss in most patients with PDR (Kohner, 1993). FFA guides the choice and extent of laser treatment in diabetic maculopathy and PDR.

An automatic identification of important events in FFA is objective and very useful both for referral and treatment. Automated analyses of FFA images for the purpose of extracting important structures as well as lesions have received some attention. Image conditioning solutions that have been proposed include illumination correction using a parametric bi-cubic model for the illumination function (Cree et al., 1999) and noise suppression for a sequence of angiogram images based on bilateral filtering (Guo et al., 2005b). In FFA segmentation, stochastic models have been proposed to segment the fovea, arteries and veins from the central (macular) view of FFAs (Simó and de Ves, 2001) and among lesions, microaneurysms have received much attention. Several techniques ranging from morphological to model-based have been proposed for microaneurysm segmentation (Fleming et al., 2006), (Hafez, 2002) and (A. M. Mendonça, 1999). An automated technique for measurement of blood flow in capillaries has been attempted from angiograms, for determining the effect of cardio-pulmonary bypass surgery (Jagoe et al., 1992). The foveal region of the retinal image

is processed to enhance the vascular structure and extract linear segments. The processed results from images taken before and after the bypass surgery are then compared (via a logical AND operation) to identify the differences. However, to our knowledge, there are no reports in the literature of any technique to detect the cause of PDR namely the presence of the CNP regions anywhere in the retina. Detecting and segmenting CNPs is the focus of this paper.

The clinical procedure to detect CNPs is a visual scan of an FFA image. In order to estimate the amount of area damaged, the scan is generally done on the composite image of the retina obtained after suitable mosaicing of several retinal segments. Such a procedure suffers from several drawbacks: the variable skills and subjectivity of the observer, which also depend on the quality of the images; a lack of precise understanding of the area of retina affected which helps in deciding the nature and extent of laser treatment. Automated image analysis techniques can be used to address these issues but there are several challenges in devising solutions for CNP segmentation. FFAs suffer from non-uniform illumination due to the eye geometry, imaging conditions and presence of other media opacity such as cataract. Inter-patient and intra-patient variablility is also possible. The former is due to different pupil dilations and the latter is due to the time of image capture after injection of fluorescein dye. Another compounding factor is that the mean grey level of CNPs as well as their shape and size are variable, with the size ranging from very small to very large (from 100 to 55000 pixels). Often, the boundaries of CNPs are not well defined because of an inhomogeneous textured background. Thus, the only visually distinguishing characteristic of a CNP is that it is relatively darker than its surround.

In this paper, we propose a novel method to extract and quantify regions of CNP based on modeling CNPs as valleys in the image surface. The algorithm for CNP segmentation is developed and its details are presented in the next section. Section 3 provides implementation details and illustrative test results of the algorithm. Finally, some discussions and conclusions are presented in the last section.

2 VALLEY BASED CNP SEGMENTATION

2.1 Modelling CNP Regions

As discussed earlier, CNP occurs when the capillary network in a region of the human retina stops functioning and does not supply blood to the corresponding areas. In FFAs, regions receiving normal blood supply appear as bright white regions since they carry a fluorescent dye and regions lacking in blood (due to abnormal supply of blood) appear as dark regions. Hence, regions of CNP appear as dull/dark lesions bounded by healthy vasculature.

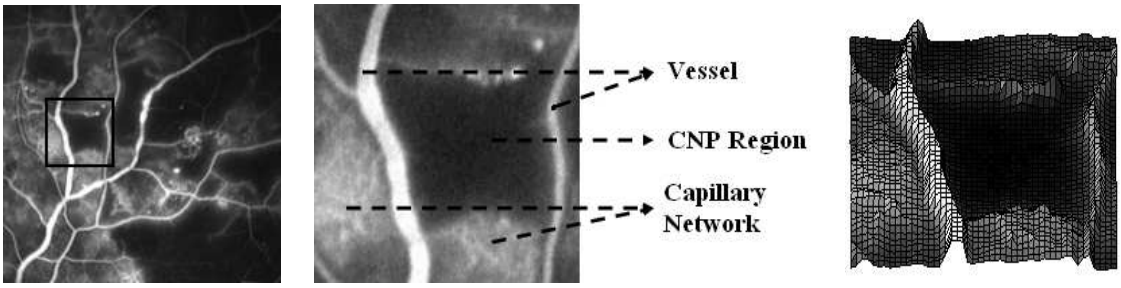
A sample FFA image and an enlarged view of a CNP region and its surroundings is shown in Fig. 1. Also, included in this figure is the surface plot of the corresponding CNP region from which we can observe that the prominent vessels, the healthy capillary network and the CNP have very different topographic charactersitics: While the major vessel appears as a ridge, the CNP appears as a valley with the healthy capillary network appearing as a plateau in the image. Hence, one can conclude that CNPs can be modelled as valleys. Watershed-based solution to valley detection (for example, (Gauch, 1999)) is possible, however, these result in oversegmentation or in the case of marker-based versions, require additional information. In the case of CNP detection, since the size of a CNP and the nature of its surround can be highly variable, obtaining such markers can be quite challenging. A better alternative is to identify the trough (lowest point on a curve) and use it to segment a CNP. Hence, we have taken a different approach to the problem and propose a technique that detects trough points and collates them across scales. We next present the details of our proposed algorithm for CNP segmentation comprising several steps.

2.2 CNP Detection Algorithm

The proposed CNP detection algorithm consists of these stages: Firstly, illumination correction (IC) is done to minimise the background intensity variation followed by denoising to eliminate noise that is frequently found in FFAs. Next, valley detection is performed to locate the seed points in the CNP regions which are used to extract the candidate CNP regions using a region growing algorithm. Finally, thresholding is done to reject false positives among the detected candidates. The processing in each of these stages are described next.

2.2.1 Illumination Correction

Nonuniform illumination is a problem in retinal colour images as well as angiograms. A camera-model based solution for illumination correction in angiograms, obtained with non-confocal imaging, is given in (Cree et al., 1999) which assumes a macula-centric view of the retina. Our images are not necessarily macula-centric and are obtained from a laser-



(a) (b) (c)

Figure 1: (a) A sample FFA image with CNP. (b) Enlarged view and (c) surface plot of the CNP region in (a).

based confocal imaging system. We modified a quotient based approach proposed for face images (Wang et al., 2004) and model the non-uniform illumination as a multiplicative degradation function which is estimated by blurring the corrupted image. Let $I(x,y)$, $I_s(x,y)$ and $I_0(x,y)$ denote the given, smoothed and corrected images, respectively and l_0 be the desired level of illumination. The corrected intensity value at location (x,y) is found as

$$I_0(x,y) = \begin{cases} I(x,y) \times \frac{l_0}{I_s(x,y)} & \text{if } I_s(x,y) < l_0 \\ I(x,y) & \text{if } I_s(x,y) \geq l_0 \end{cases} \quad (1)$$

As can be observed from Eq. 1, a pixel where the estimated illumination is greater than the ideal illumination value is not corrected. This is to ensure that the regions which are inherently bright, like the optic-disk, haemorrhages, etc., are not wrongly classified as regions of excessive illumination and corrected accordingly. When the estimated illumination value is less than the ideal illumination value, multiplication by the fraction $\frac{l_0}{I_s(x,y)}$ ensures that regions with illumination less than the l_0 are elevated to the ideal illumination value. Moreover, contrast at such a pixel is improved by a factor of $\frac{l_0}{I_s(x,y)}$ thereby removing the need for subsequent brightness and contrast operations, as required in the case of quotient-image based technique. A sample FFA image and corresponding illumination corrected image is shown in Fig. 2.

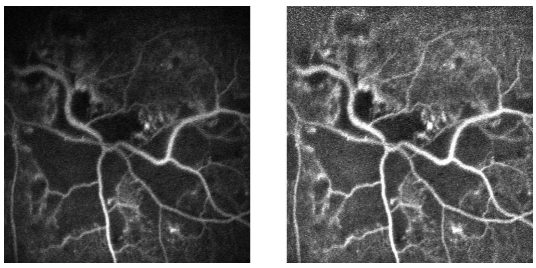


Figure 2: A sample FFA image and Illumination corrected image.

2.2.2 Noise Removal

The laser-based imaging produces fine-grain speckle type of noise in the angiograms as can be seen in Fig. 2. A bilateral filter-based approach proposed for color and gray scale images in (Tomasi and Manduchi, 1998) has been successfully applied to denoise images in an angiogram sequence (Guo et al., 2005a). The strength of bilateral filter based denoising is its ability to denoise without compromising edge quality. This is due to the filter's nonlinear characteristic which permits one to take into account the spatial distance as well the photometric similarity of a pixel to its neighbors. The spatial context is provided by a domain filter while the photometric similarity is controlled by a range filter. We use a version of the bilater filter for our noise removal task which is described next. Given an input pixel $I(P)$, the output pixel $I_0(P)$ is found as

$$I_0(P) = \frac{\sum_w I(Q) W_d(P,Q) W_r(P,Q)}{\sum_w W_d(P,Q) W_r(P,Q)} \quad (2)$$

where P and Q are position vectors, w is the current neighbourhood and W_d, W_r are Gaussian kernels of the domain and range filters respectively. The edge preservation feature of the bilateral filter can be seen in the results of preprocessing (illumination correction + denoising) in Fig.6 (b).

2.2.3 CNP Segmentation

Now we turn to the main task of detecting and segmenting CNP regions. Since we have modelled CNPs as valleys, a valley detection algorithm is needed to detect seed points in the CNP regions. As the CNPs vary widely in size, the valleys can be extended. Hence, a multiresolution approach is appropriate. The strategy we have adopted is to reduce the valleys to a single trough point via a pyramidal decomposition and then detect them using a trough detector at each

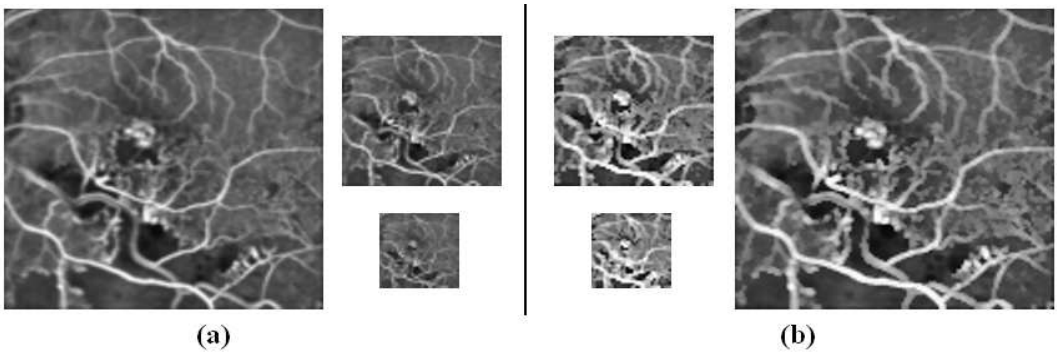


Figure 3: An FFA image at levels 3, 4, 5 in the (a) conventional and (b) proposed pyramidal decomposition.

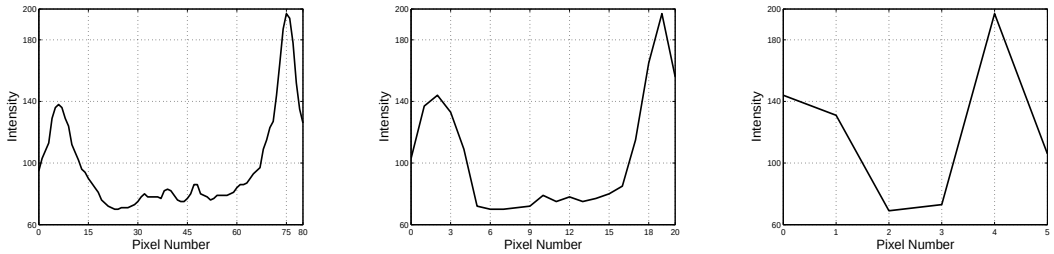


Figure 4: Intensity profile of a CNP and its surround, at levels (a) 1, (b) 3 and (c) 5 of the image pyramid.

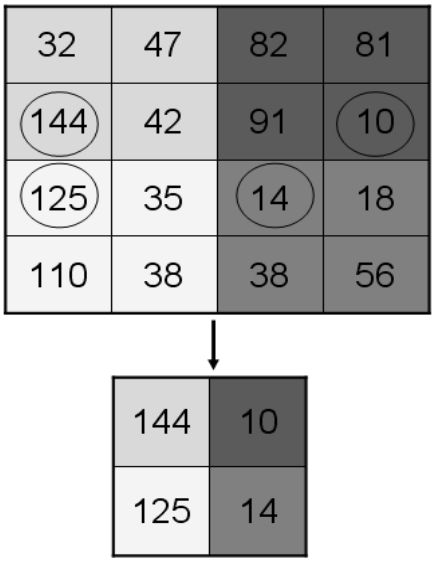


Figure 5: Example demonstrating Equation 3.

level and collating them. Each of these steps are described next.

Extrema pyramid decomposition - A conventional pyramidal decomposition based on averaging and subsampling is inadequate for the problem at hand. This can be illustrated with an example shown in Fig. 3. It can be seen that the averaging process dulls the entire image and will therefore adversely affect CNP detection based on troughs. Another drawback with the averaging process is the difficulty in localising of the trough points in the full resolution image when performing the upsampling process after trough detection. In the problem at hand, the CNP regions are generally bigger and darker relative to the brighter regions which are thin. Averaging and down-sampling will result in the bright regions to disappear faster than the CNP regions, whereas for locating troughs, it would help to more or less retain the bright regions across several levels while accepting some loss in the CNP area. Hence, to preserve the relation between a CNP and its surround, and maintain the depth of the valley across levels, we need a method for pyramidal decomposition that will minimise the CNP regions at a much faster rate compared to the brighter surrounding regions. This calls for a controlled multi-resolution technique. One option is to generate a pyramid by retaining intensity maxima which will ensure the thinner bright regions are largely preserved during downsampling. However, this is detrimental to the relative contrast between a CNP and its surround as it elevates the average intensity of the CNP regions. The end result is a lowering of the depth of

the troughs, which is undesiarbale. A better alternative is to generate the pyramid through an adaptive selection of pixels. The solution we propose is a technique for decomposition which is based on intensity extrema. Specifically, given an image I_1 of size $M \times N$ a L-level decomposition is found as follows:

$$I_l(m,n) = \left\{ \begin{array}{ll} min\{g_{i,j}(m,n)\} & \text{if } g_{i,j}(m,n) \leq t \\ max\{g_{i,j}(m,n)\} & \text{otherwise .} \end{array} \right\} \quad (3)$$

$$\forall\, i,j = 0,1$$

where $g_{i,j}(m,n) = I_{l-1}(2m+i, 2n+j)$, with $l = 2,...L$ and t is a suitable threshold, taken to be the global mean in our experiments. An illustration of the equation is given in Fig. 5 for $t = 100$.

In an extrema pyramidal decomposition of an angiogram, the CNP regions diminish in size at a much faster rate than non-CNP regions across the levels. This is illustrated in Fig. 3 (b) where the thickness of vessels are more or less preserved but the CNPs are reduced to near-dots in the lowest level image. This effect is also seen from the intensity profiles shown in Fig. 4, along a horizontal line passing through a CNP region at different levels. The width of the valley reduces from 50 pixels at the first level, to about 1 pixel in the fifth level whereas the image has been downsampled by 16 between these levels. The relative brightness value (130) of the ridge and valley regions is preserved as a result of not performing a smoothing operation.

Trough detection - A trough is defined as the lowest point on a curve. Alternatively, the brightness at a trough is a local minimum. Since the context in which CNPs, and hence troughs, occur is variable in an angiogram, two parameters can be used to characterise a trough: μ, the mean brightness of the surround and P, the peak factor which represents the depth of the trough. These two parameters are used to develop the following trough detection algorithm in which the image is denoted by $I(x,y)$.

For every pixel (x,y) do the following:

1. *Initialize a Boolean variable isTrough = False.*
2. *Check if I(x,y) is a local minimum in a $M \times M$ neighborhood.*
3. *If yes, then calculate the mean (μ) of a $N \times N$ neighborhood, with $N > M$. Else, do nothing.*
4. *Let $T = \mu * P$ and check if $I(x,y) < T$.*
5. *If yes, then isTrough = True.*
6. *If isTrough = True, then mark I(x,y) as a trough pixel. Else, do nothing.*

The threshold T represents the depth of the valley from the mean μ. Since the image pyramid retains extrema, this threshold value has to be carefully chosen to ensure that enough seed pixels are captured in a valley while minimising the possibility of false alarms. A region with low μ is likely to be a CNP region and hence the required depth for that region is less whereas the same may not be true if μ is high and hence, a stricter condition is required in the latter case. Thus, choosing T proportional to μ is appropriate. Furthermore, since trough detection is carried out at multiple levels a peak factor has to be chosen for each level. A guiding factor in this choice is that due to retention of extremas, the likelihood of the local minima being a CNP region will be higher at upper levels. Hence, the peak factor should be progressively increased with the levels in the pyramid.

After performing trough detection at all levels, the results are combined with a simple logical *OR* operation. For locating the seed pixels in the original image, the fact that the extrema of four pixels is selected at every level is used iteratively.

CNP region extraction - The detected trough points can serve as seed points for region based approach to segmenting the CNP regions. Although geometric methods can potentially yield better results, as an initial experiment we chose to use a simple region growing technique for extracting the CNP regions as it was computationally simpler. Given the variability of the appearance of the CNP regions within and across images, the traditionally used intensity-based homogeneity criterion for region growing is not suitable. Instead, by noting that CNP regions are smooth, the better alternative is to perform the pixel aggregation in the variance space.

In our experiments, the range for the variance was taken to be ± 4. In order to reject false candidates, a final thresholding operation was performed. A threshold based on the global mean intensity was applied since the global mean is always lowered with the presence of CNPs.

3 IMPLEMENTATION AND RESULTS

The proposed algorithm was implemented as follows. In the illumination correction stage, the ideal illumination l_0 in (1) was set to be roughly half the maximum grey value in the image or 120. The smoothed image was obtained with a 30×30 Gaussian mask on a subsampled (by 4) version of the original image for faster processing. The resulting image was upsampled, one level at a time, while blurring the image

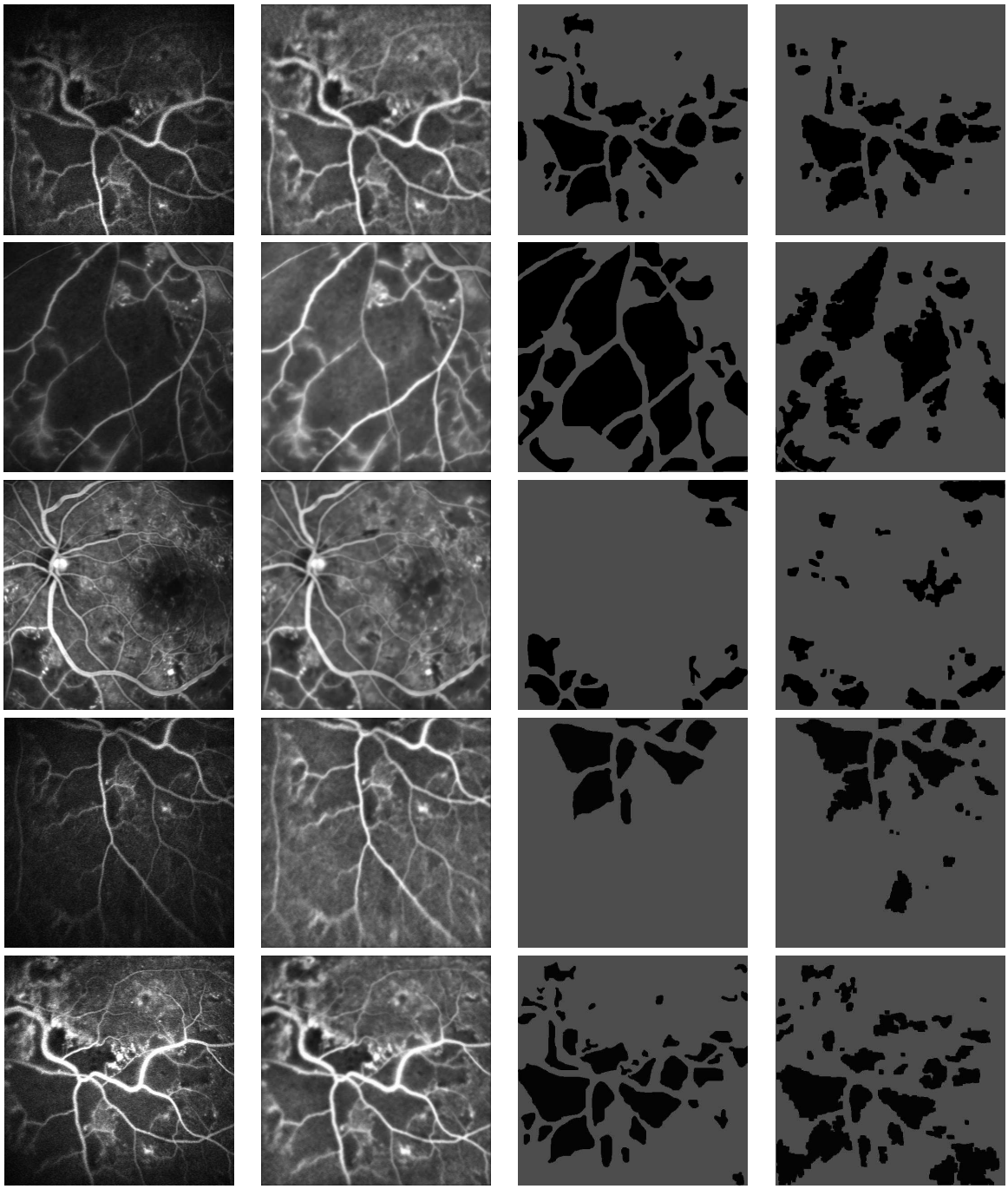


Figure 6: Results: (a) Sample images, (b) preprocessed image, (c) corresponding ground truths and (c) segmented results with CNP regions shown in black.

using a 5×5 Gaussian mask at each upsampled level. For denoising, a filter kernel size of 9×9 was used and σ for the domain and range filters were fixed at 3 and 10 respectively. For valley detection, a 5-level pyramid was generated; M, N were fixed at 5 and 7 respectively and the peak factor was incremented by 0.02 at each level in the pyramid. In region growing, the variance was calculated over a 5×5 neighbourhood.

The proposed CNP segmentation algorithm was

tested on 40 images which contained many CNPs. These were acquired from the digital confocal scanning laser ophthalmoscope of Heidelberg Retina Angiograph. The images were of retinal segments for which the ground truth, in the form of boundaries of CNPs, were prepared manually by a retina expert (a co-author). Some sample test images along with corresponding ground truth and results of our CNP segmentation algorithm, with a peak factor of 0.41, are shown in Fig. 6 (a), (b) and (c) respectively. CNP regions are shown in black in both ground truth and segmented results. The five sample test images indicate the variability in images in terms of quality, size of CNPs and presence of other structures such as optic disk, macula and microaneurysms. A quantitative assessment of the algorithm was done using a ROC curve and not a FROC curve since the area of CNP is of clinical interest. A comparison between computed and marked CNP segments was done on a pixel by pixel basis. By using the peak factor as a control parameter, the obtained ROC curve, shown in Fig. 7, was found to have an area under the curve (AUC) of 0.842 and a distance (D_i) to the ideal point (1,0) of 0.35. The ideal values for AUC and D_i are 1 and 0 respectively.

4 CONCLUDING REMARKS

An unsupervised algorithm for automatically segmenting CNPs from FFA images has been presented. Its overall performance is quite good as indicated by the ROC curve and the AUC, D_i metrics. Since there is no reported work on this problem it is not possible to do any benchmarking. A visual inspection of segmented results indicates that the algorithm successfully detects CNPs of all sizes, however, it tends to undersegment large CNPs because the IC stage intensifies the variability within CNPs. A failure analysis indicates that the macula region gets mislabled as a CNP (as seen in the bottom row of Fig. 6) since the two have similar characteristics, and CNPs in the image periphery tend to be missed since the valley model is weak in this region.

The main focus of the presented work was on detection of CNPs. Hence, while the performance our current implementation is quite good, there is scope for improvement of the algorithm's performance: geometric techniques such as fast marching method in (Malladi and Sethian, 2006) can be used to more accurately extract the CNP region boundaries while a pixel-based classifier will help improve the rejection of the false alarms. Likewise, incorporation of a macula detection stage will help the algorithm from falsely classifying macula as a CNP region.

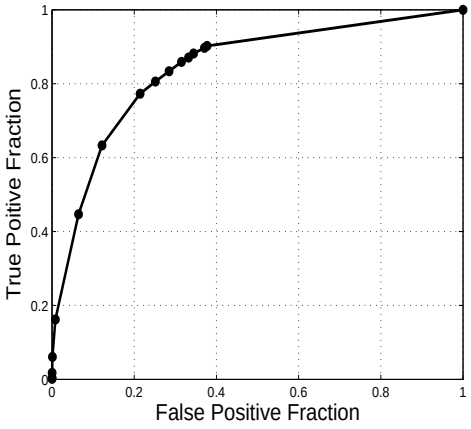


Figure 7: ROC plot.

Finally, it should be noted that the ground truth generation process for CNPs is a laborious one. Retina experts found it challenging to draw precise boundaries because they often appear to be intricate or ill-defined. This points to the need for the use of multiple expert-markings for a fuller evaluation of the algorithm. Such an exercise might also shed light on the degree of observer bias in CNP segmentation.

REFERENCES

A. M. Mendonça, A. J. Campilho, J. M. N. (1999). Automatic segmentation of microaneurysms in retinal angiograms of diabetic patients. In *Proc. International Conf. on Image Analysis and Processing*, pages 728 – 733.

Cree, M. J., Olson, J. A., C.McHardy, K., Sharp, P. F., and Forresters, J. V. (1999). The preprocessing of retinal images for the detection of fluorescein leakage. *Phys. Med. Biol.*, 44:293 – 308.

Fleming, A. D., Philip, S., Goatman, K. A., Olson, J. A., and Sharp, P. F. (2006). Automated microaneurysm detection using local contrast normalization and local vessel detection. In *IEEE Trans. on Medical Imaging*, volume 25, pages 1223 – 1232.

Gauch, J. M. (1999). Image segmentation and analysis via multiscale gradient watershed hierarchies. In *IEEE Trans. on Image Processing*, volume 8, pages 69 – 79.

Guo, X.-X., Lu, Y.-N., Xu, Z.-W., Liu, Z.-H., Wang, Y.-X., and Pang, Y.-J. (2005a). Noise suppression of fluorescein angiogram sequences using bilateral filter. In *Proc. International Conf. on Machine Learning and Cybernetics*, volume 9, pages 5366 – 5371.

Guo, X.-X., Lu, Y.-N., Xu, Z.-W., and Pang, Y.-J. (2005b). An eawa filter for denoising of filtering of fluorescein angiogram sequences. In *Proc. International Conf. on*

Computer and Information Technology, pages 614 – 618.

Hafez, M. (2002). Using adaptive edge technique for detecting microaneurysms in fluorescein angiograms of the ocular fundus. *Proc. Mediterranean Electrotechnical Conf.*, pages 479 – 483.

Jagoe, R., Arnold, J., Blauth, C., Smith, P., K.M.Taylor, and Wootton, R. (1992). Measurement of capillary dropout in retinal angiograms by computerised image analysis. *Pattern Recognition Letters*, 13:143 – 151.

Kohner, E. M. (1993). Diabetic retinopathy. *BMJ*, 307(6913):1195 – 1199.

Malladi, R. and Sethian, J. A. (2006). Fast methods for shape extraction in medical and biomedical imaging. In *R Malladi (ed) Geometric Methods in Bio-medical Image Processing. Berlin, Springer*, pages 49–61.

Simó, A. and de Ves, E. (2001). Segmentation of macular fluorescein angiographies. a statistical approach. *Pattern Recognition*, 34:795 – 809.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proc. International Conf. on Computer Vision*, pages 839–846.

Wang, H., Li, S. Z., and Wang, Y. (2004). Generalized quotient image. In *Proc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 498 – 505.

AN EFFICIENT METHOD FOR VESSEL WIDTH MEASUREMENT ON COLOR RETINAL IMAGES

Alauddin Bhuiyan, Baikunth Nath, Joselito Chua and Kotagiri Ramamohanarao
Department of Computer Science and Software Engineering and NICTA Victoria Research Laboratory
The University of Melbourne, Australia
{*bhuiyanm, bnath, jjchua, rao*} *@csse.unimelb.edu.au*

Keywords: Microvascular Sign, Gradient Operator, Adaptive Region Growing Technique, Texture Classification, Gabor Energy Filter Bank, Fuzzy C-Means Clustering.

Abstract: Vessel diameter is an important factor for indicating retinal microvascular signs. In automated retinal image analysis, the measurement of vascular width is a complicated process as most of the vessels are few pixels wide. In this paper, we propose a new technique to measure the retinal blood vessel diameter which can be used to detect arteriolar narrowing, arteriovenous (AV) nicking, branching coefficients, etc. to diagnose related diseases. First, we apply the Adaptive Region Growing (ARG) segmentation technique to obtain the edges of the blood vessels. Following that we apply the unsupervised texture classification method to segment the blood vessels from where we obtain the vessel centreline. Then we utilize the edge image and vessel centreline image to obtain the potential pixels pairs which pass through a centreline pixel. We apply a rotational invariant mask to search the pixel pairs from the edge image. From those pixels we calculate the shortest distance pair which will be the vessel width for that cross-section. We evaluate our technique with manually measured width for different vessels' cross-sectional area and achieve an average accuracy of 95.8%.

1 INTRODUCTION

Accurate measurement of retinal vessel diameter is an important part in the diagnosis of many diseases. A variety of morphological changes occur to retinal vessels in different disease conditions. The change in width of retinal vessels within the fundus image is believed to be indicative of the risk level of diabetic retinopathy; venous beading (unusual variations in diameter along a vein) is one of the most powerful predictor of proliferate diabetic retinopathy. Generalized and focal retinal arteriolar narrowing and arteriovenous nicking have been shown to be strongly associated with current and past hypertension reflecting the transient and persistent structural effects of elevated blood pressure on the retinal vascular network. In addition, retinal arteriolar bifurcation diameter exponents have been shown to change significantly in patients with peripheral vascular disease and arteriosclerosis and a variety of retinal microvascular abnormalities have been shown to relate to the risk of stroke (Lowell et al., 2004). Therefore, an accurate measurement of vessel diameter and geometry is necessary for effective diagnosis of such diseases.

The measurement of the vascular diameter is critical and a challenging task whose accuracy depends on the accuracy of the segmentation method. A review for the blood vessel segmentation is provided in the literature (Bhuiyan et al., 2007a). The study of vessel diameter measurement is still an open area for improvement. Zhou et al. (Zhou et al., 1994) have applied a model-based approach for tracking and to estimating widths of retinal vessels. Their model assumes that image intensity as a function of distance across the vessel displays a single Gaussian form. However, high resolution fundus photographs often display a central light reflex (Brinchman-hansan and Heier, 1986). Intensity distribution curves is not always of single Gaussian form, so that using a single Gaussian model for simulating intensity profile of vessel could produce poor fits and subsequently provide inaccurate diameter estimations (Gao et al., 2001). Gao et al. (Gao et al., 2001) model the intensity profiles over vessel cross section using twin Gaussian functions to acquire vessel width. This technique may produce poor results in case of minor vessels where the contrast is less. Lowell et al. (Lowell et al., 2004) have proposed an algorithm based on fitting a local 2D vessel model, which can measure vascular width to an accuracy of about one third of

a pixel. However, the technique is biased on smooth data (image) and suffers from measuring the width of minor vessels where the contrast is very less.

In this paper, we introduce a new algorithm, based on vessel centreline and edges information. We apply the adaptive region growing technique to segment the vessels edges (Bhuiyan et al., 2007a) and the unsupervised texture classification method to segment the vessels and detect the centreline (Bhuiyan et al., 2007b). For each selected centerline pixel we map the edge image of the retinal vessels edge pixels and find all the potential line end points or pairing pixels on opposite edge passing through this centreline pixels. From these potential lines we find the line that has the minimum length and consider this as the vessel width for that cross-sectional area. In this way, we can measure the width of the blood vessel continuing through the centreline of all the vessels. A specific feature of our technique is that it can calculate the vessel width when it is one pixel wide.

The rest of the paper is organized as follows: Section 2 introduces the proposed method of blood vessel width measurement. Edge based blood vessel segmentation technique is described in section 3. Section 4 illustrates the vessel centreline detection procedure. The vessel width measurement method is described in section 5. The experimental results are provided in section 6 and finally the conclusion and future research directions are drawn in section 7

2 PROPOSED METHOD

We propose the blood vessels' width measurement algorithm based on the vessel edge and centreline. The major advantage of our technique is that it is less sensitive to noise and work equally for the low contrast vessels (particularly for minor vessels). We adopt two segmented images that are produced from the original RGB image. At first, we apply the ARG segmentation technique to obtain the vessel edges, then we apply the unsupervised texture classification method to segment the blood vessels from where we obtain the vessel centreline. We map the vessel centreline image and pick any of the vessel centreline pixel. For that particular pixel we apply a rotational invariant mask whose centre is that pixels position and searches the potential pixels from the edge image using a continuous increment of lower to higher distance and orientation. For each case, if the gray scale value of that pixel position is 255 or white it finds the mirror of this pixel by searching through a fixed angle (exactly incrementing 180 degree) but in variable distance. This is to give the flexibility and consistency to our method as the centreline pixels may not be in the exact position of vessel centre. In this way, we can obtain all the potential pairs (line end points) which pass through that centreline pixel. From those pairs we calculate the minimum distance/length pair which is the width of that cross-section of the blood vessel. Figure 1 depicts the overall technique of our proposed method.

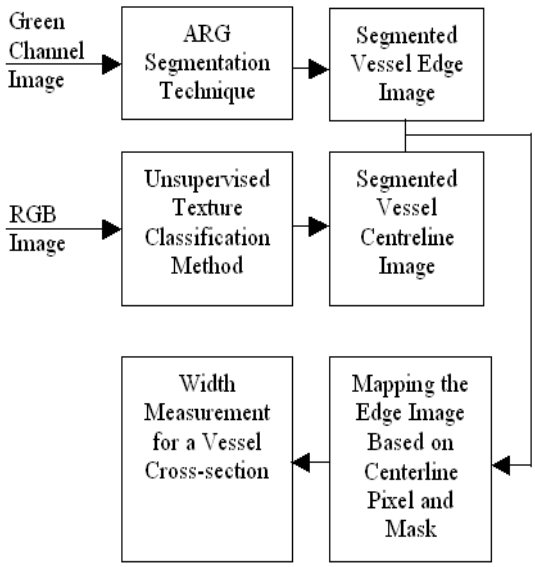


Figure 1: The overall system for measuring blood vessel width.

3 VESSEL EDGE DETECTION

We implemented the vessel segmentation technique based on vessel edges. In the following subsections we provide a brief illustration of this method.

3.1 Preprocessing of Retinal Image

Adaptive Histogram Equalization (AHE) method is implemented, using MATLAB, to enhance the contrast of the image intensity by transforming the values using contrast-limited adaptive histogram equalization (Figure 2).

3.2 Image Conversion

The enhanced retinal image is converted into gradient image (Figure 2) using first order partial differential operator. The gradient of an image $f(x,y)$ at location (x,y) is defined as the two dimensional vector (Gon-

zalez and Wintz 1987)

$$G[f(x,y)] = [G_x, G_y] = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \quad (1)$$

For edge detection, we are interested in the magnitude $G[f(x,y)]$ and direction $\alpha(x,y)$of the vector, generally referred to simply as the gradient and denoted and commonly takes the value of

$$\begin{aligned} G[f(x,y)] &\approx |G_x| + |G_y| \\ \alpha(x,y) &= \tan^{-1}(G_y/G_x) \end{aligned} \quad (2)$$

where the angle is measured with respect to the x axis.

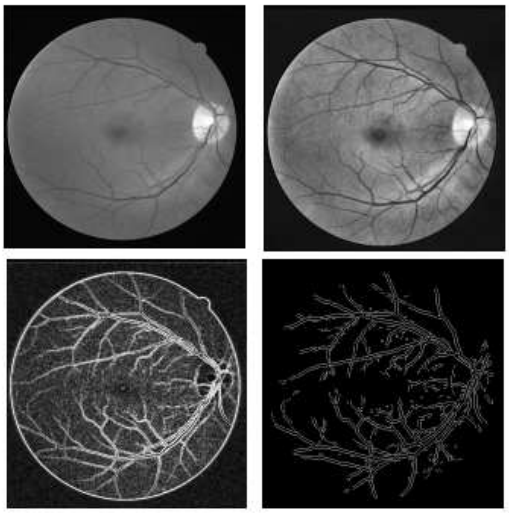


Figure 2: Original retinal image, its Adaptive Histogram Equalized image (top; left to right), the Gradient Image and final ARG output image (bottom; left to right).

3.3 Adaptive Region Growing Technique

The edges of vessels are segmented using region growing procedure (Gonzalez et al., 2004) that groups pixels or sub regions into larger regions based on gradient magnitude. As the gradient magnitude is not constant for the whole vessel we need to consider an adaptive gradient value that gradually increases or decreases to append the pixel to a region. We call it an adaptive procedure, as the difference of neighboring pixels intensity value is always adapted for the region growing process. The region growing process starts with appending the pixels that pass certain threshold value. For region growing we find the intensity difference between a pixel belonging to a region and its neighboring potential region growing pixels. The pixel is considered for appending in that region if the difference is less than a threshold value. The threshold value is calculated by considering the maximum differential gradient magnitude for any neighboring pixels with equal (approximately) gradient direction. Region growing should stop when no more pixels satisfy the criteria for inclusion in that region. In the region growing process each region is labeled with a unique number. For that purpose we construct a cell array with region number and its pixel position. The image is scanned in a row-wise manner until its end, and each pixel that satisfies our criteria is taken into account for growing a region with its 8-neighborhood connectivity.

3.4 Parallel Region Detection

We calculate the parallel edges (regions) by considering pixel orientation belonging to each region. At first, we pick the region number and belonging pixel coordinates from the constructed cell array. Then we grouped the region/regions parallel to each region, which is calculated by mapping the pixels gradient direction. For each region every pixel is searched from its potential parallel region and once a maximum number of pixels match with the other region we consider it as parallel to that region. We consider all regions and once a region is considered we assigned a flag value to that region so that it will not be considered again. In this way we can only filter the vessels from the region and discard all other regions, which are background noise or other objects like haemorrhage, macula, etc in the retinal image.

3.5 Experimental Results

We considered DRIVE database (DRIVE-database, 2004) and applied our technique on five images for initial assessment. For performance evaluation we employed an expert to find the number of vessels in the original image and detected output image (Figure 2). We achieved an overall 94.98% detection accuracy.

4 VESSEL CENTRELINE DETECTION

We implemented the unsupervised texture classification based vessel segmentation method from which we detect the vessel centreline. We consider Gaussian and $L^*a^*b^*$ perceptually uniform color spaces with the original RGB image for texture feature extraction. To extract features, a bank of Gabor energy filters with three wavelengths and twenty-four orientations is applied in each selected color channel. Then

a texture image is constructed from the maximum response of all orientations for a particular wavelength in each color channel. From the texture images, a feature vector is constructed for each pixel. These feature vectors are classified using the Fuzzy C-Means (FCM) clustering algorithm. Finally, we segment the image based on the cluster centroid value.

4.1 Color Space Transformation and Preprocessing

Generally image data is given in RGB space (because of the availability of data produced by the camera apparatus). The definition of $L^*a^*b^*$ is based on an intermediate system, known as the CIE XYZ space (ITU-Rec 709). This space is derived from RGB as below (Wyszecki and Stiles, 1982)

$$\begin{aligned} X &= 0.412453R + 0.357580G + 0.180423B \\ Y &= 0.212671R + 0.715160G + 0.072169B \\ Z &= 0.019334R + 0.119193G + 0.950227B \end{aligned} \tag{3}$$

$L^*a^*b^*$ color space is defined as follows:

$$\begin{aligned} L^* &= 116 f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n)] - f(Z/Z_n) \end{aligned} \tag{4}$$

where $f(q) = q^{1/3}$ if $q < 0.008856$ and is constant 7.87+16/116 otherwise. X_n, Y_n and Z_n represent a reference white as defined by a CIE standard illuminant, D_{65} in this case. This is obtained by setting $R = G = B = 100$ in (1), $q \in \{X/X_n, Y/Y_n, Z/Z_n\}$.

Gaussian color model can also be well approximated by the RGB values. The first three components $\hat{E}$, $\hat{E_\lambda}$ and $\hat{E_{\lambda\lambda}}$ of the Gaussian color model (Taylor expansion of the Gaussian weighted spectral energy distribution at Gaussian central wavelength and scale) can be approximated from the CIE 1964 XYZ basis when taking $\lambda_0 = 520nm$ (Gaussian central wavelength) and $\sigma_\lambda = 55nm$ (scale) as follows (Geusebroek et al., 2001)

$$\begin{pmatrix} \hat{E} \\ \hat{E_\lambda} \\ \hat{E_{\lambda\lambda}} \end{pmatrix} = \begin{pmatrix} -0.48 & 1.2 & 0.28 \\ 0.48 & -0.4 & -0.4 \\ 1.18 & -1.3 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{5}$$

The product of (3) and (5) gives the desired implementation of the Gaussian color model in RGB terms (Figure 3). The Adaptive Histogram Equalization method was implemented, using MATLAB, to enhance the contrast of the image intensity.

4.2 Texture Feature Extraction

Texture generally describes second order property of surfaces and scenes, measured over image intensities.

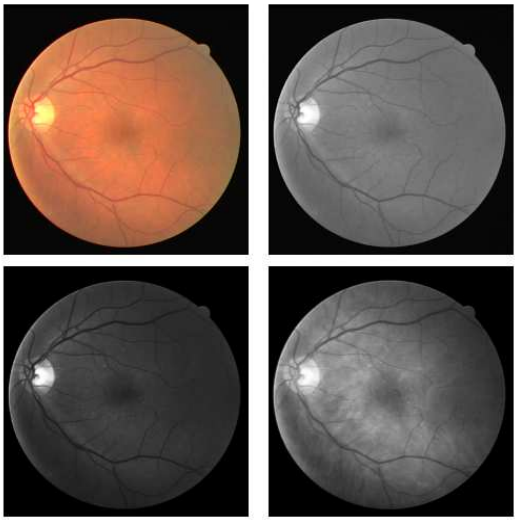


Figure 3: Original RGB and its Green channel image (top; left to right), Gaussian transformed first and second component image (bottom; left to right).

A Gabor filter has weak responses along all orientations on the smooth (background) surface. On the other hand, when it positioned on a linear pattern object (like a vessel) the Gabor filter produces relatively large differences in its responses when the orientation parameter changes (Wu et al., 2006). Hence, the use of Gabor filters to analyze the texture of the retinal images is very promising. In the following two subsections we illustrate the Gabor filter based texture analysis method.

4.2.1 Gabor Filter

An input image $I(x,y)$, $(x,y) \in \Omega$ where Ω is the set of image points, is convolved with a 2D Gabor function $g(x,y)$, $(x,y) \in \omega$, to obtain a Gabor feature image $r(x,y)$ (Gabor filter response) as follows (Kruizinga and Petkov, 1999)

$$r(x,y) = \iint_{\Omega} I(\xi,\eta) g(x-\xi, y-\eta) d\xi d\eta \tag{6}$$

We use the following family of 2D Gabor functions to model the spatial summation properties of an image (Kruizinga and Petkov, 1999)

$$\begin{aligned} g_{\xi,\eta,\lambda,\Theta,\phi}(x,y) &= \exp\left(-\tfrac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\tfrac{x'}{\lambda}+\phi\right) \\ x' &= (x-\xi)\cos\Theta - (y-\eta)\sin\Theta \\ y' &= (x-\xi)\cos\Theta - (y-\eta)\sin\Theta \end{aligned} \tag{7}$$

where the arguments x and y specify the position of a light impulse in the visual field and $\xi, \eta, \sigma, \gamma, \lambda, \Theta, \phi$ are parameters. The pair (ξ, η) specifies the center of a receptive field in image coordinates. The standard

deviation σ of the Gaussian factor determines the size of the receptive filed. Its eccentricity is determined by the parameter γ called the spatial aspect ratio. The parameter λ is the wavelength of the cosine factor which determines the preferred spatial frequency $\frac{1}{\lambda}$ of the receptive field function $g_{\xi,\eta,\lambda,\Theta,\phi}(x,y)$. The parameter Θ specifies the orientation of the normal to the parallel excitatory and inhibitory stripe zones - this normal is the axis x' in (5). Finally, the parameter $\phi \in (-\pi,\pi)$, which is a phase offset argument of the harmonic factor $\cos(2\pi\frac{x'}{\lambda}+\phi)$, determines the symmetry of the function $g_{\xi,\eta,\lambda,\Theta,\phi}(x,y)$.

4.2.2 Gabor Energy Features

A set of textures was obtained based on the use of Gabor filters (6) according to a multichannel filtering scheme. For this purpose, each image was filtered with a set of Gabor filters with different preferred orientation, spatial frequencies and phases. The filter results of the phase pairs were combined, yielding the Gabor energy quantity (Kruizinga and Petkov, 1999):

$$E_{\xi,\eta,\Theta,\lambda} = \sqrt{r^2_{\xi,\eta,\Theta,\lambda,0} + r^2_{\xi,\eta,\Theta,\lambda,\pi/2}} \qquad (8)$$

where $r^2_{\xi,\eta,\Theta,\lambda,0}$ and $r^2_{\xi,\eta,\Theta,\lambda,\pi/2}$ are the outputs of the symmetric and antisymmetric filters. We used Gabor energy filters with twenty-four equidistant preferred orientations ($\Theta = 0,15,30,..,345$) and three preferred spatial frequencies ($\lambda = 6,7,8$). In this way an appropriate coverage was performed of the spatial frequency domain.

We considered the maximum response value per pixel on each color channel to reduce the feature vector length and complexity of training on data for the classifier. In addition, we constructed an image (Figure 4) on each color channel which was used for histogram analysis to determine the cluster number. From these images we constructed twelve element length feature vector for each pixel in each retinal image to classify them into vessel and non-vessel using the FCM clustering algorithm.

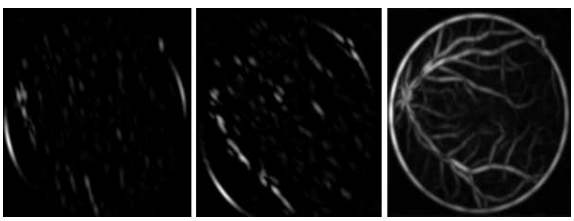


Figure 4: Texture analyzed image with the orientations of 15, 45 degrees and maximum response of all twenty-four orientations (left to right).

4.3 Texture Classification and Image Segmentation

The FCM is a data clustering technique where in each data point belongs to a cluster to some degree that is specified by a membership grade. Let $X = x_1,x_2,,x_N$ where $x \in R^N$ present a given set of feature data. The objective of the FCM clustering algorithm is to minimize the Fuzzy C-Means cost function formulated as (Bezdek, 1981)

$$J(U,V) = \sum_{j=1}^{C}\sum_{i=1}^{N}(\mu_{ij})^m ||x_i - v_j||^2 \qquad (9)$$

$V = \{v_1,v_2,,v_C\}$ are the cluster centers. $U = (\mu_{ij})_{N\times C}$ is fuzzy partition matrix, in which each member is between the data vector x_i and the cluster j. The values of matrix U should satisfy the following conditions:

$$\mu_{ij} \in [0,1], i = 1,..,N, j = 1,..,C \qquad (10)$$

$$\mu_{ij} = 1, i = 1,..,N \qquad (11)$$

The exponent $m \in [1,\infty]$ is the weighting exponent, which determines the fuzziness of the clusters. The most commonly used distance norm is the Euclidean distance $d_{ij} = ||x_i - v_j||$.

We used the Matlab Fuzzy Logic Toolbox for clustering 253440 vectors (the size of the retinal image is 512x495) in length twelve for each retinal image. In each retinal image clustering procedure, the number of clusters was assigned after analyzing the histogram of the texture image. The parameter values used for the FCM clustering were as follows. The exponent value of 2 for the partition matrix, maximum number of iterations was set to 1000 for the stopping criterion and the minimum amount of improvement being 0.00001. We received the membership values on each cluster for every vector, from which we picked the cluster number that belonged to the highest membership value for each vector and converted it into a 2D matrix. From this matrix we produced the binary image considering the cluster central intensity value which identifies the blood vessels only.

4.4 Experimental Results

Using the DRIVE database (DRIVE-database, 2004) we applied our method on five images for vessel segmentation. For performance evaluation, we detected the vessel centerline in our output segmented images and hand-labeled ground truth segmented (GT) images applying the morphological thinning operation (Figure 5). We achieved an overall 84.37% sensitivity ($TP/(TP+FN)$) and 99.61% specificity ($TN/(TN+FP)$) where TP, TN, FP and FN are true

positive, true negative, false positive and false negative respectively. Hoover et al. (Hoover et al., 2000) method on the same five segmented images provided in 68.23% sensitivity and 98.06% specificity. Clearly, our method produces superior results.

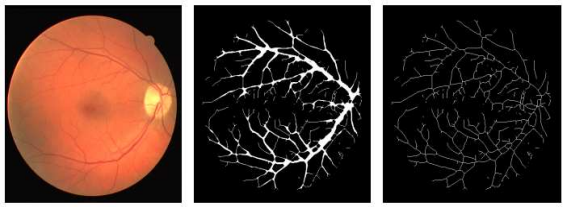


Figure 5: Original RGB image, vessel segmented image, and its centreline image (from left to right).

5 VESSEL WIDTH MEASUREMENT

After obtaining the vessels edge image and centreline image, we mapped these images to find the vessel width for a particular vessel centreline pixel position. To do this we first pick a pixel from the vessel centreline image, then we apply a mask considering this centreline pixel as its centre. The purpose of this mask is to find the potential edge pixels (which may fall in width or cross section of the vessels) in any side of that centreline pixel position. Therefore, we will apply the mask to the edge image only. For searching all the pixel positions inside the mask, we calculate the pixel position by shifting by one up to the size of the mask and rotating each position from 0 to 180 degrees at the same time. For increasing the rotation angle we use the step size (depending on the size of the mask) less then 180/(mask length). Therefore, we can access every cell in the mask using this angle.

For each obtained position we search the edge image gray scale value to check whether it is an edge pixel or not. Once we find an edge pixel we then find it's mirror by shifting the angle of 180 degree and increasing the distance from one to the maximum size of the mask (Figure 6). In this way we produce a rotational invariant mask and pick all the potential pixel pairs to find the width or diameter of that cross sectional area.

$$\begin{aligned} x1 &= x' + r * \cos\theta \\ y1 &= y' + r * \sin\theta \end{aligned} \tag{12}$$

where (x',y') is the vessel centreline pixel position, r=1,2,..(mask size)/2 and $\theta = 0,..,180^o$. For any pixel position, if the gray scale value in the edge image is 255 (white or edge pixel) then we find the pixel $(x2,y2)$ in the opposite edge (mirror of this pixel) considering $\theta = (\theta + 180)$ and varying r.

After applying this operation we obtain the pairs of pixels which are on the opposite edges (at line end points) giving imaginary lines passing through the centerline pixels (Figure 6). From these pixels pairs we find the minimum Euclidian distance $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$, the width of that crosssection. In this way, we can measure the width for all vessels including the vessels' with one pixel wide (for which we have the edge and the centreline itself).

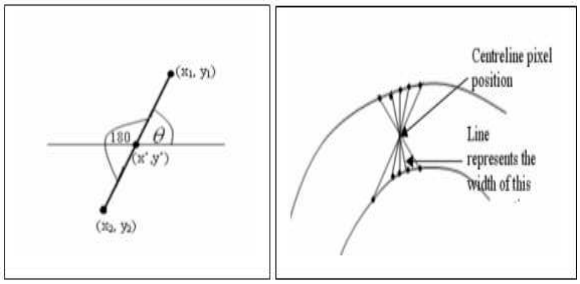


Figure 6: Finding the mirror of an edge pixel(left) and width or minimum distance from potential pairs of pixels (right).

6 EXPERIMENTAL RESULTS AND DISCUSSION

We used the centreline images and edge images for measuring the width of the blood vessels. We measure the accuracy qualitatively by comparing with the width measured by plotting the centreline pixel and its surround edge pixels. We considered ten different vessel cross-sections of these images and observed that our method is working very accurately. Figure 7 portrays the Grid for a cross-section of a blood vessel where c is the centreline pixel and w1 to w8 are potential width end points. Figure 8 depicts the detected width for some cross-sectional points indicating in white lines (enlarged).

	251	252	253	254	255	256	257	258	259	260	261	262	263
455				W8	W7	W6							
456							W5						
457								W4					
458									W3	W2			
459	W1					C					W1		
460		W2											
461			W3	W4									
462					W5	W6							
463							W7						
464								W8					

Figure 7: Grid showing the potential width edge pairs for a cross-section with centreline pixel C.

For quantitative evaluation we considered ten images (each 3072×2048 which captured with the

Table 1: Measuring the accuracy of the automatic width measurement.

Cross-section	Centreline pixel X_c	Y_c	Detected width end points X_1	Y_1	X_2	Y_2	Auto. width (A)	Accuracy (%)	Error (%)
1	2055	629	2068	632	2046	628	22.361	99.14	0.86
2	1859	519	1871	519	1850	520	21.024	97.50	2.50
3	2259	815	2259	811	2259	824	13	99.46	0.54
4	2350	1077	2350	1070	2350	1084	14	87.61	12.39
5	2233	1317	2239	1314	2239	1322	11.314	93.49	6.51
6	2180	1435	2189	1431	2172	1440	19.235	95.39	4.61
7	2045	1451	2055	1452	2042	1452	13	85.55	14.45
8	1683	1500	1691	1509	1680	1496	17.029	87.52	12.48
9	1579	617	608	1593	630	1573	23.409	98.48	1.52
10	1434	855	853	1436	859	1432	7.211	85.48	14.52
11	1443	1000	999	1446	1004	1440	7.81025	91.23	8.77
12	1618	1331	1335	1623	1330	1617	7.81025	89.54	10.46
13	1475	1164	1169	1479	1162	1474	8.6023	83.20	16.80

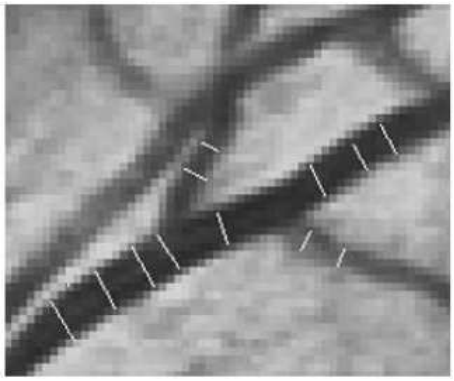


Figure 8: Measured vessel width showing by the white lines in an image portion.

Canon D-60 digital fundus camera) with manually measured width on different cross-sections from Eye and Ear Hospital, Victoria, Australia. For each cross-section, we received the graded width by five different experts who are trained retinal vessel graders of that institution. For manual grading a computer program was used where the graders could zoom in and out at will, moving around the image and selecting various parts. We applied our technique on these images to produce the edge image and vessel centreline image. We considered these images and randomly picked ninety-six cross-sections of vessels varying width from one to twenty-seven pixels. We measured the width for each cross-section by our automatic width measurement technique (we call it automatic width, A) and considered the five manually measured width (we call it manual width) by experts. We calculated the average of the manual width (μ), the standard deviation on manual widths (σ_m) and considered the following formula to find the error,

$$\begin{aligned} E &= \left| \frac{\frac{(\mu-\sigma_m)-A}{(\mu-\sigma_m)} + \frac{(\mu+\sigma_m)-A}{(\mu+\sigma_m)}}{2} \right| \\ &= \left| 1 - \frac{\mu \times A}{\mu^2 - \sigma_m^2} \right| \end{aligned} \tag{13}$$

In equation (13), we considered $(\mu \pm \sigma_m)$ to normalize it. This formula is a good measure as the error rate will be less if it is within the interval one standard deviation. With this formula, we calculated the error and accuracy in all ninety-six cross-section and achieve an average of 95.8% accuracy (maximum accuracy is 99.58% and minimum accuracy is 83.20%) in the detection of vessel width. We found the maximum error is 16.80% which is 2.04 pixel and the minimum error is 0.698% which is 0.139 pixel. Table 1 and 2 depict the manual and automatic width measurement accuracy on different cross-sections in an image. We compared our technique with (Lowell et al., 2004) which achieved the maximum accuracy of 99% (did not mention the average accuracy for all cross-sections) with minimum pixel error of 0.34. Using the same formula, $|(\mu - A)/\mu|$), we achieved 100% accuracy. Clearly, our technique is performing better.

7 CONCLUSIONS AND FUTURE WORK

In this paper we proposed a new and efficient technique for blood vessels width measurement. This approach is a robust estimator of vessel width in the presence of low contrast and noise. The results obtained are promising and the detected width can be used to measure different parameters (nicking, nar-

Table 2: Manually measured widths for an image cross-sections.

Cross-section	Manually measured width (in Micron) One	Two	Three	Four	Five	Mean width (μ) (in pixel)	Standard Deviation (σ_m)
1	112.42	117.53	107.31	117.53	112.42	22.2	0.8366
2	107.31	112.42	107.31	117.53	107.31	21.6	0.8944
3	66.43	76.65	61.32	71.54	61.32	13.2	1.3088
4	61.32	71.54	61.32	71.54	56.21	12.6	1.3416
5	56.21	66.43	56.21	66.43	66.43	12.2	1.0954
6	107.31	107.31	102.2	102.2	97.09	20.2	0.8366
7	56.21	66.43	45.99	61.32	66.43	11.6	1.6733
8	86.87	107.31	102.2	107.31	97.09	19.6	1.6733
9	132.86	127.75	112.42	132.86	107.31	24	2.3452
10	45.99	51.1	35.77	56.21	35.77	8.8	1.7889
11	40.88	56.21	35.77	45.99	45.99	8.8	1.4832
12	35.77	51.1	45.99	56.21	40.88	9	1.5811
13	35.77	45.99	35.77	45.99	30.66	7.6	1.3416

rowing, branching coefficients, etc.) for diagnosing various diseases. Currently, we are working on the blood vessels' bifurcation and cross-over detection where the measured width is contributing as an important information for perceptual grouping process.

ACKNOWLEDGEMENTS

We would like to thank David Griffiths (Research Assistant, The University of Melbourne and Eye and Ear Hospital, Melbourne, Australia) for providing us with the manually measured width images and data.

REFERENCES

Bezdek, J. (1981). Pattern recognition with fuzzy objective function algorithms. *Plenum Press, USA*.

Bhuiyan, A., Nath, B., and Chua, J. (2007a). An adaptive region growing segmentation for blood vessel detection from retinal images. *Second International Conference on Computer Vision Theory and Applications*, pages 404–409.

Bhuiyan, A., Nath, B., Chua, J., and Kotagiri, R. (2007b). Blood vessel segmentation from color retinal images using unsupervised classification. *In the proceedings of the IEEE International Conference of Image Processing*.

Brinchman-hansan, O. and Heier, H. (1986). Theoritical relations between light streak characterstics and optical properties of retinal vessels. *Acta Ophthalmologica*, 179(33).

DRIVE-database (2004). http://www.isi.uu.nl/research/databases/drive/, image sciences institute, university medical center utrecht, the netherlands.

Gao, X., Bharath, A., Stanton, A., Hughes, A., Chapman, N., and Thom, S. (2001). Measurement of vessel diameters on retinal images for cardiovascular studies. *Proceedings of Medical Image Understanding and Analysis*, pages 1–4.

Geusebroek, J., Boomgaard, R. V. D., Smeulders, A. W. M., and Geerts, H. (2001). Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):1338–1350.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). Digital image processing using matlab. *Prentice Hall*.

Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210.

Kruizinga, P. and Petkov, N. (1999). Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing*, 8(10):1395–1407.

Lowell, J., Hunter, A., Steel, D., Basu, A., Ryder, R., and Kennedy, R. L. (2004). Measurement of retinal vessel widths from fundus images based on 2-d modeling. *IEEE Transactions on Medical Imaging*, 23(10):1196–1204.

Wu, D., Zhang, M., and Liu, J. (2006). On the adaptive detetcion of blood vessels in retinal images. *IEEE Transactions on Biomedical Engineering*, 53(2):341–343.

Wyszecki, G. W. and Stiles, S. W. (1982). Color science: Concepts and methods, quantitative data and formulas. *New York, Wiley*.

Zhou, L., Rzeszotarsk, M. S., Singerman, L. J., and Chokreff, J. M. (1994). The detection and quantification of retinopathy using digital angiograms. *IEEE Transactions on Medical Imaging*, 13.

IDENTIFICATION OF TIME-VARYING T-WAVE ALTERNANS FROM 20-MINUTE ECG RECORDINGS

Issues Related to TWA Magnitude Threshold and Length of ECG Time Series

Laura Burattini
Department of Electromagnetics and Bioengineering, Polytechnic University of Marche, via Brecce Bianche 60131 Ancona, Italy
l.burattini@univpm.it

Wojciech Zareba
Heart Research Follow-Up Program, Cardiology Unit, Department of Medicine and Department of Biomedical Engineering, University of Rochester, 601 Elmwood Ave, Rochester, NY 14642-8679, USA
wojciech_zareba@urmc.rochester.edu

Roberto Burattini
Department of Electromagnetics and Bioengineering, Polytechnic University of Marche, via Brecce Bianche 60131 Ancona, Italy
r.burattini@univpm.it

Keywords: Signal processing of the digital electrocardiographic signal, Repolarization variability, T-wave alternans.

Abstract: Aim of this study was the assessment of a T-wave alternans (TWA) identification procedure based on application of an adaptive match filter (AMF) method, recently developed by ourselves, to a 20-minute digital ECG recording (ECG20). Three-lead ECG20 tracings from 35 patients who survived an acute myocardial infarction (AMI-group) and 35 healthy subjects (H-group) were analysed. The AMI-group showed, on average, increased levels of TWA ($P<0.01$). Considering that noise may cause false positive TWA detection, a threshold (THR_{TWA}) was defined for TWA magnitude (TWAM) as the mean TWAM +2SD over the H-group. TWAM exceeding this threshold identified a TWA-positive (TWA+) subject as one at increased risk of sudden cardiac death. Fifteen (43%) AMI-patients vs. zero H-subjects were detected as TWA+. This result meets clinical expectation. TWA manifested as a non stationary phenomenon that could even be missed in all TWA+ subjects if our AMF (as well as any other technique) was applied to a single short-term 128-beat ECG series, as usually done in previous reports. In conclusion, our AMF-based TWA identification technique, applied to 20-minute ECG recordings, yields a good compromise between reliability of time-varying TWA identification and computational efforts.

1 INTRODUCTION

T-wave alternans (TWA) is an electrophysiological phenomenon which consists of two-to-one beat-to-beat changes in the morphology (amplitude, shape and, sometimes, polarity) of the electrocardiografic (ECG) T wave. According to the literature, visible and non-visible (microvolt) forms of TWA in ECG recordings play an important role in the arrhythmogenesis of failing myocardium (Schwartz and Malliani, 1975; Zareba et al., 1994; Adam et al., 1984; Smith et al., 1988; Rosembaum et al., 1994; Kusmirek and Gold, 2007; Klingenheben and Ptaszynski, 2007; Narayan, 2007). Visible forms of TWA are infrequent. Non-visible TWA requires computerized analysis of digital ECG recordings to be recognized and parameterized in terms of amplitude and duration. Thus, in the effort to assess a clinically useful marker of sudden cardiac death, development of methods for non-invasive automatic detection of microvolt TWA has been a major challenge in the last two decades (Rosenbaum et al., 1996; Klingenheben et al., 2000; Ikeda et al., 2002; Tapanainen et al., 2001; Bigger and Bloomfield, 2007; Ikeda et al., 2006). Factors that may prevent a reliable TWA quantification must be controlled by

signal preprocessing, such as high frequency noise filtering, detection of R peaks, RR stability testing, and removal of baseline deviation from the isoelectric line.

Recently, we developed a new adapting match filter (AMF; Burattini et al., 2006) method for automatic TWA detection, which, differently from other reported techniques, does not require any pre-processing of the ECG tracing, with the only exception of R-peak detection. Making use of simulated (Burattini et al., 2006) and experimental (Burattini et al., 2007) data, we showed that this method yields an improvement in reliability of TWA detection over a previously reported correlation method (Burattini,1998; Burattini et al., 1999).

Like any other TWA detection technique, our AMF needs to be applied to ECG tracings with no significant heart-rate variability and with a low noise level. As a consequence, these techniques have traditionally been applied to short-term ECG series, typically consisting of 128 consecutive heart beats. This rises the issue as to whether 128 beats portray sufficient information on the presence of TWA. To address this issue, in the present study we analyzed 3-lead (X,Y,Z) 20-minute digital ECG recordings (ECG20). Our goal was to demonstrate that repeated applications of our AMF-based method to several tracings of 128 heart beats, within an ECG20, yields a good compromise between reliability of TWA identification and computational efforts. Our analysis was performed on Holter ECG recordings from patients who survived a myocardial infarction since these are known to show increased levels of TWA, compared to healthy subjects (Ikeda et al., 2002; Pelicano et al., 2006; Ikeda et al., 2000; Puletti et al., 1980).

2 METHODS

2.1 Clinical Data

Our study involved 35 healthy subjects (H-group; RR=0.93±0.17 s) and 35 patients who survived an acute myocardial infarction (AMI-group; RR=0.88±0.14 s). For a better traceability during the analysis, healthy subjects were identified as H01, H02, … H35. Analogously, AMI patients were identified as AMI01, AMI02, … AMI35.

A twenty-minute, three-lead (X,Y,Z) digital Holter recording was obtained from each individual in resting conditions, making use of Burdick recorders (Burdick Inc., Milton, WI). Sampling rate was 200 samples per sec. Series of 128 consecutive cardiac beats were extracted every 10 seconds from each tracing. Each series underwent our TWA identification procedure as described below. Because extraction of 128 cardiac beats every 10 s causes a data overlap (on average 109 s for the H-group, and 103 s for the AMI-group) between two consecutive series, effects of this overlapping were tested vs. an extraction procedure (data selection every 128 beats) that avoids data overlapping.

2.2 T-Wave Alternans Detection by Adaptive Match Filter

Our adaptive match filter method (AMF), specifically designed to detect TWA (Burattini et al., 2006 and 2007), was applied to each ECG series of 128 heart beats.

To avoid cases where TWA could be driven by heart-rate variability (Adam et al., 1984; Rosembaum et al., 1994; Burattini, 1998; Burattini et al., 1999), an ECG time series has to be characterized by a stable heart rate to be eligible for TWA analysis. Specifically, we required that:

$$SDRR < 0.1 \cdot MRR \quad (1)$$

where MRR and SDRR are mean and standard deviation of RR intervals (in s).

Under this condition, the TWA phenomenon is assumed to be characterized by a specific frequency of half heart rate: f_{TWA}=0.5 cycles per beat, or f_{TWA}= 1/(2*MRR) Hz. To account for physiological variations of the RR interval, a narrow frequency band, instead of a single frequency, was assumed here to characterize the TWA phenomenon. On this basis, our AMF was designed as a passband filter with its passing band centred in f_{TWA}. Technically, the AMF was implemented as a 6^{th} order bidirectional Butterworth band-pass filter, having the passing band $2 \cdot df_{TWA}$=0.12 Hz wide (value experimentally found) and centred at a frequency that adapts to mean RR interval. In particular, our AMF was designed as a cascade of a low pass filter (LPF) with cut-off frequency $f_{LPF}= f_{TWA}+ df_{TWA}$, and a high pass filter (HPF) with a cut-off frequency $f_{HPF}= f_{TWA}-df_{TWA}$. The squared module of the AMF transfer function is expressed by the following equation:

$$\left|H_{AMF}(w)\right|^2 = \left|H_{LPF}(w)\right|^2 \cdot \left|H_{HPF}(w)\right|^2 = \\ = \frac{1}{1+\left(\frac{w}{w_{LPF}}\right)^{2n}} \cdot \frac{\left(\frac{w}{w_{HPF}}\right)^{2n}}{1+\left(\frac{w}{w_{HPF}}\right)^{2n}} \quad (2)$$

were n=3 (half of AMF order), $w_{LPF}=2\pi f_{LPF}$, and $w_{HPF}=2\pi f_{HPF}$. Being the AMF applied in a bidirectional fashion, no phase delay occurs. Thus, the AMF is expected to detect the TWA signal by filtering out not only noise and baseline wandering, but also any other ECG component but the TWA.

The TWA signal provided by the AMF is a time domain, constant phase and, possibly, amplitude-modulated sinusoid with its maxima and minima over the T-waves. A local estimate of TWA amplitude (A_{TWA}), associated to each single beat, is directly given by the sinusoid amplitude in correspondence of the T-wave apexes. If the T wave of a beat is alternating, its A_{TWA} is greater than zero. In our procedure, all local A_{TWA} values are used to compute global (i.e. relative to all 128 beats of the ECG series) estimates of TWA characteristic parameters. In particular, the following global parameters were determined: TWA duration (TWAD, beat; defined as the total number of beats with alternating T-waves), TWA amplitude (TWAA, μV; defined as the mean A_{TWA} over all alternating T-waves), and TWA magnitude (TWAM, beat·μV; defined as the product of TWAA times TWAD). TWAM is used to detect the presence of TWA, since it includes information about both TWAA and TWAD. Moreover, TWAM allows identification of different TWA episodes (such as those short in time and high in amplitude, or long in time and low in amplitude), which would not be detected if only TWAD or TWAA, respectively, were used. Thus, the AMF allows characterization of non-stationary (i.e. time varying) characteristics of the TWA signal, when present.

TWAD, TWAA and TWAM parameter values are determined in each available lead. Corresponding values from the three different leads (X,Y,Z) are then averaged for final TWA characterization relative to a specific 128-beat series. The series with the highest TWAM is assumed as the most representative of the entire 20-minute recording.

2.3 Identification of TWA-Positive Subjects

Considering that noise and artefacts may be detected as TWA episodes, once TWA is identified and parameterized, there is a need to define the TWA level that characterizes a TWA-positive subject as one at increased risk of sudden cardiac death. Taking advantage of the H-group involved in our study, the mean+2SD value of the TWAM distribution over this group was assumed as the normality threshold (THR_{TWA}) of TWA magnitude. Thus, subjects with TWAM greater than THR_{TWA} were considered as TWA positive (TWA+).

2.4 Statistical Analysis

Lilliefors test (Lilliefors, 1967), was used to evaluate the hypothesis that each data vector or parameter vector had a normal distribution (significance was set at 5% level) and could be expressed as mean ± SD. Comparisons between two groups of normally distributed samples were performed with two-tailed, non-paired Student's t-test (statistically significant difference was assumed at $P<0.05$).

3 RESULTS

Application of our AMF method to an entire 20-minute ECG recording (ECG20), with 128 beat ECG series selected every 10 s, yielded normally distributed TWA parameters with mean±SD for H-group and AMI-group as given in Table 1. The AMI-group was found to be characterized by having significantly higher TWAD, TWAA, and TWAM. The threshold value (THR_{TWA}), as defined in Methods, was 4176 beat·μV (that is, 2730+2×723). With this threshold, fifteen patients of the AMI-group (i.e. 43%) were classified as TWA+. No subject of the H-group showed relevant TWA.

Extraction of ECG time series every 128 beats (no overlap between two consecutive series) provided a lower number of TWA+ among AMI-patients (eleven cases, i.e. 31%), and significantly lower estimates of TWA duration, amplitude and magnitude parameters, with respect to the extraction procedure performed every 10 s (Table 2).

Application of our AMF method to 128-beat series taken in proximity of minutes 0 (t_0), 5 (t_5), 10 (t_{10}), 15 (t_{15}), and 20 (t_{20}), yielded even lower numbers of TWA+ patients associated with significant reduction of mean TWAM, compared to ECG20 with ECG time series extracted every 10 s (Table 3). Eight (23%) out of the 15 (43%) patients identified as TWA+ when analyzing ECG20 (namely, AMI02, AMI10, AMI11, AMI15, AMI18, AMI19, AMI22 and AMI24), were never detected as TWA+ when using single 128-beat series. Four (11%) patients were detected as TWA+ at time t_0 and t_{15}, two (6%) at t_{10}, and only one (3%) at t_5 and t_{20}.

A representative example of the time course of TWAD, TWAA and TWAM parameters averaged over the three leads in our AMI01 patient is

displayed in Fig. 1a to c. Panel c clearly shows that TWAM, which has been assumed as a marker of the presence of TWA, crosses the THR_{TWA} value at different time instants. Under-threshold values of TWAM are due to a simultaneous decrement of both TWAD and TWAA (panels a and b).

Because TWAM shows fluctuations with threshold-crossing within ECG20, TWA could even be missed in all TWA+ subjects if our AMF was applied to a single, short-term 128-beat ECG series. Confirmation of this statement is found in Fig. 2, where TWAM waves from all fifteen TWA+ patients are displayed. Arrow pointers in proximity of the eleventh minute mark 128-beat ECG series with under-threshold TWAM which would miss all TWA+ cases.

4 DISCUSSION

To satisfy the requirement of heart-rate stability for reliable TWA detection, short-term ECG series have been considered for TWA identification in most reported studies. Indeed, spectral analysis has been the first technique proposed in the literature for automatically detecting TWA (Adam et al., 1984). ECG series of 128-consecutive beats were considered for its application because this is the minimal requirement to guarantee reliable spectral analysis. Since then, ECG time series of 128-consecutive beats have been traditionally used for TWA quantification. Thus, the issue arises as to the reliability of using a single 128 beat sequence to detect TWA+ cases. The present study was designed to address this issue by applying our AMF-based method for TWA detection (Burattini et al., 2006). Comparison was performed among the results obtained from 128 beat ECG series selected 1) every 10 s (data overlap), 2) every 128 beats (no data overlap), 3) every 5 minutes, in a time frame of 20 minutes. For this technical investigation, we considered a population of 35 AMI-patients compared with a population of 35 H-subjects. The H-population was used as reference to define a threshold (THR_{TWA}) for TWAM parameter provided by our method as a marker to identify a remarkable level of TWA.

A novel finding of our analysis was that, based on the defined threshold, the use of a unique 128 beat ECG series is unsuitable to unmask and detect TWA. An explanation of this shortcoming is found in that TWA is a transient phenomenon characterized by time-varying TWAD, TWAA and TWAM parameters (Fig. 1). As shown in Fig. 2, under-threshold values of TWAM, assumed as marker of TWA, would miss TWA+ patients if a unique 128-beat ECG series in proximity of the eleventh minute was used.

Table 1: Comparison between TWA duration (TWAD), amplitude (TWAA), and magnitude (TWAM) distributions (mean±SD) in the H-group and AMI-group. Data refer to 20-minute ECG recordings (ECG20) with 128 beat time series extracted every 10 s.

	H-group	AMI-group	t-test
TWAD (beat)	75±13	87±11	P<0.01
TWAA (μV)	43±14	56±22	P<0.01
TWAM (beat*μV)	2730±723	3982±1386	P<0.001

Table 2: Comparison between TWA duration (TWAD), amplitude (TWAA), and magnitude (TWAM) distributions (mean±SD) in AMI-group. Data refer to 20-minute ECG recordings with 128 beat time series extracted every 10 s (ECG20 overlap) or every 128 beats (ECG20 no overlap).

AMI-group	ECG20 overlap	ECG20 no overlap	t-test
TWA+	15	11	
TWAD (beat)	87±11	81±11	P<0.001
TWAA (μV)	56±22	50±21	P<0.001
TWAM (beat*μV)	3982±1386	3453±1253	P<0.001

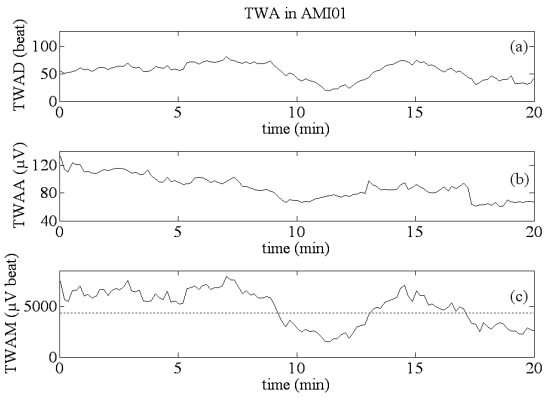


Figure 1: TWA in the AMI01 patient. Panels a, b, and c: respectively, TWA duration (TWAD), TWA amplitude (TWAA), and TWA magnitude (TWAM) as functions of time. In panel c the normality threshold is represented with a dotted line.

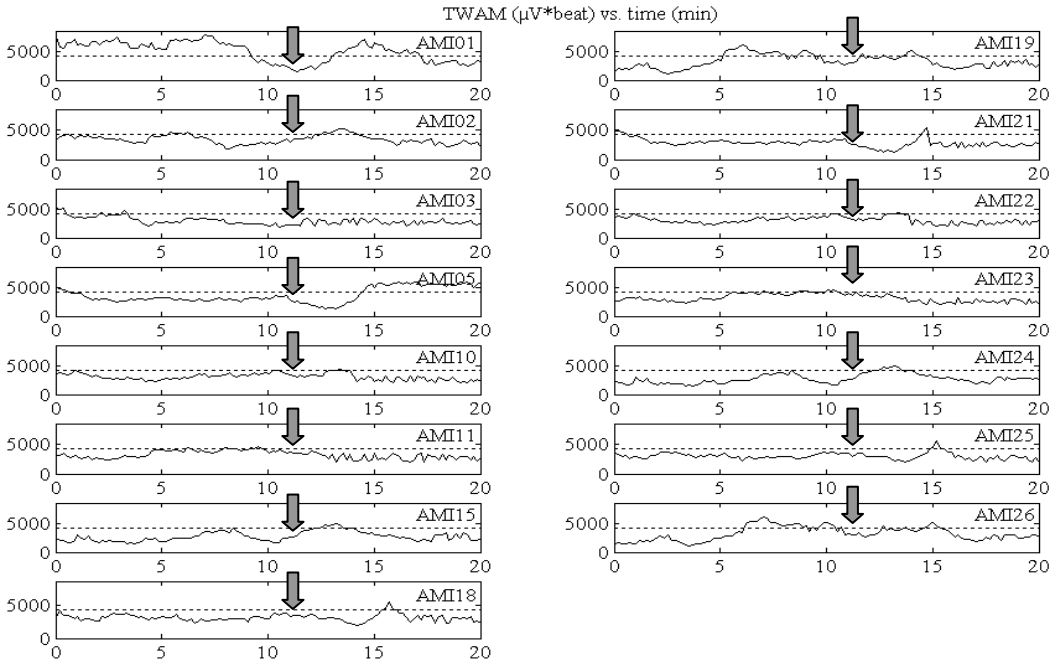


Figure 2: TWAM waves from our TWA+ patients. All them would not be recognised as TWA+ if a single 128 beat series about the eleventh minute (arrows) was used, since TWAM is under threshold (dot line) about this time.

Table 3: TWA+ patients of AMI-group identified by our AMF method applied to the entire 20-minute ECG recording with 128 beats series extracted every 10 s (ECG 20 overlap), and to a single 128-beat series taken in proximity of minutes 0 (t_0), 5 (t_5), 10 (t_{10}), 15 (t_{15}), and 20 (t_{20}). TWAM: TWA magnitude; TWA+: TWA positive patient. Student's t-test is used to compare the mean TWAM value over each considered 128-beat series (t_0, t_5, ..,t_{20}) with mean TWAM over ECG 20.

AMI-group	ECG 20 overlap	t_0	t_5	t_{10}	t_{15}	t_{20}
TWA+ individual patients	AMI01	AMI01	AMI1	-	AMI01	-
	AMI02	-	-	-	-	-
	AMI03	AMI03	-	-	-	-
	AMI05	AMI05	-	-	AMI05	AMI05
	AMI10	-	-	-	-	-
	AMI11	-	-	-	-	-
	AMI15	-	-	-	-	-
	AMI18	-	-	-	-	-
	AMI19	-	-	-	-	-
	AMI21	AMI21	-	-	-	-
	AMI22	-	-	-	-	-
	AMI23	-	-	AMI23	-	-
	AMI24	-	-	-	-	-
	AMI25	-	-	-	AMI25	-
	AMI26	-	-	AMI26	AMI26	-
Total TWA+	15 (43%)	4 (11%)	1 (3%)	2 (6%)	4 (11%)	1 (3%)
TWAM (beat*μV)	3982±1386	2708±1324	2352±974	2330±1030	2494±1298	2397±1178
t-test		P<0.001	P<0.001	P<0.001	P<0.001	P<0.001

We identified in 3-lead (X, Y, Z) 20-minute digital ECG recordings a time frame that achieves a good compromise between reliability of TWA identification and computational efforts. Moreover, a 20-minute time frame is short enough to be possibly obtained in controlled conditions, so that noise and heart-rate variability due to emotional or physical factors may not interfere significantly.

The definition of a threshold for TWAM is a critical issue. The value identified here as mean TWAM+2SD over our H-group yielded no presence of TWA in this population, since the under-threshold level of TWA is considered as background noise. Our finding of no TWA in H-subjects is consistent with what is commonly recognised in clinics. Nevertheless, further studies on populations of clinical relevance are desirable to define an optimal normality threshold.

Several techniques have been proposed in the literature for TWA detection (Adam et al., 1984; Nearing et al., 1991; Burattini et al., 1999; Burattini et al., 2006). Among these, the spectral method, pioneered by Adam et al. (1984), is the most widely used in clinics. However, being TWA a transient (i.e. non-stationary) phenomenon (Kusmirek and Gold, 2007; Cox et al., 2007; Richter et al., 2005; and present study), a time-domain approach, as our AMF method, appears, from a theoretical point of view, more appropriate since it provides local (i.e. relative to the single beat) as well as global (i.e. relative to the entire ECG series under analysis) TWA parameterization. Moreover, it is able to discriminate between TWA phenomena sustained-in-time (minutes) but low-in-amplitude and short-in-time (few beats) but large-in-amplitude. Because these two different kinds of TWA could potentially have different clinical implications (statement to be confirmed by future clinical studies), it appears worthwhile to have a TWA detection method, which allows discrimination between them. Such a discrimination is not allowed by the spectral method, which works, by definition, under the hypothesis of stationary signal, and provides TWA measurements that are averaged over the entire ECG time series under analysis (128 beats). As a consequence, no local (at the beat level) parameterization is possible with the spectral method.

Two more TWA detection techniques proposed in the literature are the correlation method (Burattini et al., 1999) and the complex demodulation (Nearing et al., 1991), which operate in the time domain. Compared to the correlation method, our AMF improves TWA detection in the presence of baseline wanderings (Burattini et al., 2006). Complex demodulation is computationally very heavy and has never been used for practical purposes. In addition, compared to any other TWA detection algorithm, our AMF does not require pre-processing of the ECG tracing, because noise and ECG frequency components other than f_{TWA}, are simultaneously filtered out.

Our study suggests to analyse 20 minute ECG recordings by applying our AMF to 128 beat ECG time series selected every 10 seconds within a 20-minute time frame. As a consequence, the TWA global parameters (TWA duration, amplitude and magnitude) associated to a time instant are the result of an integration procedure over a 128 beats window, corresponding (see Methods), on average, to 119 s for the H-group, and 113 s for the AMI-group. This, of course, results in a significant overlap of data sets. A certain degree of overlap, however, is necessary. In fact TWA episodes could be divided into shorter ones during the windowing procedure for 128 ECG time series extraction, and this operation could prevent a correct TWA detection and quantification. In addition, to be eligible for TWA analysis, a 128 beat ECG is required to satisfy the heart-rate stability condition (eq. 1). The presence of local arrhythmic or noise conditions, including ventricular premature beats, artefacts, as well as false-positive and false-negative beat detections, may cause rejection of a 128 beat ECG. If no overlap among ECG time series is present, all information on TWA in the time frame belonging to the rejected ECG series is lost. Rather, if a certain degree of overlap is allowed, some information on TWA can be recovered from a close ECG time series not affected by the local noise factor. The time resolution recovering, consisting of the transition from the global (relative to the entire ECG series) domain to the local (relative to the single beat) domain, is possible only with time-domain TWA detection methods, through the availability local TWA amplitude measure (A_{TWA}).

Computational efforts limit the frequency of time series extraction from an ECG20. Results of the present study indicate that application of our AMF-based method to 128 beat series extracted every 10 s is a good compromise between reliability of non stationary, transient TWA identification and computational efforts. This kind of analysis, in fact, can be routinely performed in real time in a clinic or doctor's office using a standard personal computer.

REFERENCES

Schwartz P.J., Malliani A., 1975. Electrical alternation of T wave: clinical and experimental evidence of its relationship with the sympathetic nervous system and with the Long QT Syndrome. *Am Heart J*,; 89(1): 45-50.

Zareba W., Moss A.J., Le Cessie S., Hall W.J., 1994. T wave Alternans in idiopathic Long QT Syndrome. *J Am Coll Cardiol,* 23(7): 1541-1546.

Adam D.R., Smith J.M., Akselrod S., Nyberg S., Powell A.O., Cohen R.J., 1984. Fluctuations in T-wave morphology and susceptibility to ventricular fibrillation. *J Electrocardiol,* 17(3): 209-218.

Smith J.M., Clancy E.A., Valeri C.R., Ruskin J.N., Cohen R.J., 1988. Electrical alternans and cardiac electrical instability. *Circulation,* 77: 110-121.

Rosenbaum D.S., Jackson L.E., Smith J.M., Garan H., Ruskin J.N., Cohen R.J., 1994. Electrical alternans and vulnerability to ventricular arrhythmias. *N Engl J Med,* 330(4): 235-241.

Kusmirek S.L., Gold M.R., 2007. Dynamic changes of T-wave alternans: does it predict short-term arrhythmia vulnerability?. *J Cardiovasc Electrophysiol,* 18(5): 518-519.

Klingenheben T., Ptaszynski P., 2007. Clinical significance of microvolt T-wave alternans. *Herzschrittmacherther Elektrophysiol,* 18(1):39-44.

Narayan S.M., 2007. T-wave alternans and human ventricular arrhythmias: what is the link?. *J Am Coll Cardiol,* 49(3): 347-349.

Rosenbaum D.S., Albrecht P., Cohen R.J., 1996. Predicting sudden cardiac death from T wave alternans from the surface electrocardiogram: promises and pitfalls. *J Cardiovasc Electrophysiol,* 7(11): 1095-1111.

Klingenheben T., Zabel M., D'Agostino R.B., Cohen R.J., Hohnloser S.H., 2000. Predictive value of T-wave alternans for arrhythmic events in patients with congestive heart failure. *Lancet,* 356(9230): 651-652.

Ikeda T., Saito H., Tanno K., Shimizu H., Watanabe J., Ohnishi Y., Kasamaki Y., Ozawa Y., 2002. T-wave alternans as predictor for sudden death after myocardial infarction. *Am J Cardiol,* 89(1): 79-82.

Tapanainen J.M., Still A.M., Airaksinen K.E., Huikuri H.V., 2001. Prognostic significance of risk stratifiers of mortality, including T wave alternans, after acute myocardial infarction: results of prospective follow-up study. *J Cardiovasc Electrophysiol,* 12(6): 645-652.

Bigger J.T., Bloomfield D.M., 2007. Microvolt T-wave alternans: an effective approach to risk stratification in ischemic cardiomyopathy?. *Nat Clin Pract Cardiovasc Med,* 4(6):300-301.

Ikeda T., Yoshino H., Sugi K., Tanno K., Shimizu H., Watanabe J., Kasamaki Y., Yoshida A., Kato T., 2006. Predictive value of microvolt T-wave alternans for sudden cardiac death in patients with preserved cardiac function after acute myocardial infarction: results of a collaborative cohort study. *J Am Coll Cardiol,* 48(11): 2268:2274.

Burattini L., Zareba W., Burattini R., 2006. Automatic detection of microvolt T-wave alternans in Holter recordings: Effect of baseline wandering. *Biomedical Signal Processing and Control,* 1(2): 162-168.

Burattini L., Zareba W., Burattini R., 2007. Heart-rate adapting match filter detection of T-wave alternans in experimental Holter ECG recordings. *In: Proceeding of the Fifth IASTED International Conference on Biomedical Engineering*, Gardner J.W. Ed., ACTA Press, Zurich: 346-351.

Burattini L., 1998. *Electrocardigraphic T-wave alternans detection and significance.* Doctoral thesis. University of Rochester, Rochester NY.

Burattini L., Zareba W., Moss A.J., 1999. Correlation method for detection of transient T-wave alternans in digital ECG recordings. *Ann Noninvasive Electrocardiol,* 4(4): 416-424.

Pelicano N., Oliveira M., da Silva N., Anao A., Feliciano J., Fiarresga A., Alves S., Silva S., Quininha J., 2006. Assessment of T-wave alternans after acute myocardial infarction: influence of timing of PTCA on cardiac electrical stabilization. *Rev Port Cardiol,* 25(4): 379-387.

Ikeda T., Sakata T., Takami M., Kondo N., Tezuka N., Nakae T., Noro M., Enjoji Y., Abe R., Sugi K., Yamaguchi T., 2000. Combined assessment of T-wave alternans and late potentials used to predict arrhythmic events after myocardial infarction. A prospective study. *J Am Coll Cardiol,* 35(3): 722-730.

Puletti M., Curione M., Righetti G., Jacobellis G., 1980. Alternans of the ST segment and T wave in acute myocardial infarction. *J Electrocardiol,* 13(3): 279-300.

Lilliefors H.W., 1967. On the Kolmogorov-Smirnov test for normality with men and variance unknown. *J Am Stat Assoc,* 62:399-402.

Cox V., Patel M., Kim J., Liu T., Sivaraman G., Narayan S.M., 2007. Predicting arrhythmia-free survival using spectral and modified-moving average analyses of T-wave alternans. *Pacin Clin Electrophysiol,* 30(3):352-358.

Richter S., Duray G., Hohnloser S.H., 2005. How to analyze T-wave alternans. *Heart Rhythm*, 2(11):1268-1271.

Nearing B.D., Huang A.H., Verrier R.L., 1991. Dynamic tracking of cardiac vulnerability by complex demodulation of the T wave. *Science,* 252(5004): 437-440.

POSTERS

TRADITIONAL AVERAGING, WEIGHTED AVERAGING, AND ERPSUB FOR ERP DENOISING IN EEG DATA

A Comparison of the Convergence Properties

Andriy Ivannikov, Tommi Kärkkäinen, Tapani Ristaniemi
Department of Mathematical Information Technology, University of Jyväskylä
P.O. Box 35 (Agora), FIN-40014, Jyväskylä, Finland
andriy.ivannikov@cc.jyu.fi, tommi.karkkainen@mit.jyu.fi, tapani.ristaniemi@mit.jyu.fi

Heikki Lyytinen
Department of Psychology, University of Jyväskylä
P.O. Box 35 (Agora), FIN-40014, Jyväskylä, Finland
heikki.lyytinen@psyka.jyu.fi

Keywords: Convergence, EEG, ERP, Denoising, Weighted Averaging, SNR, Source Separation.

Abstract: In this article we compare the convergence rates at increase of the number of processed trials of the three methods applied nowadays in electroencephalography research to denoising of event-related potentials: traditional averaging, weighted averaging, and ERPSUB. We derive the weighted averaging procedure by maximizing signal-to-noise ratio in the averaged subject responses and show, thereby, that maximizing signal-to-noise ratio criterion is equivalent to minimizing the originally proposed mean-square error criterion in the sense of the weighted averaging problem solving. Moreover, in order to characterize fully the performance of the selected methods, we compare also noise reduction rates in estimates of event-related potentials provided by methods, while the number of processed trials increases.

1 INTRODUCTION

Reliable characterization of event-related potentials (ERPs) is a central task in electroencephalography (EEG) data processing. ERP is a concept used in EEG research to denote brain electromagnetic potentials occurring as responses to the external or mental events, whose quantitative understanding underlies many neuropsychological studies and clinical diagnosis (Huttunen et al., 2007; Luu et al., 2004; Makeig et al., 1999; Näätänen, 1992). However, the signal-to-noise ratio (SNR) is very low in a single measurement (trial) of the brain response following the stimulation event, which makes it impossible to identify ERP characteristics, such as amplitude and latency, reliably. In order to increase SNR and, hence, estimate reliably ERP characteristics, many trials of equal length and synchronized to the same event are measured from different locations on the scalp (channels) and averaged channel-wise (see Sect. 2). Averages of many trials for every channel are assumed to have high SNR and important ERP characteristics can be identified then from the averages with the accuracy depending on the number of trials used for averaging.

Moreover, besides improving the reliability of the estimates of ERP characteristics, it is also important to shorten the experiment time, because subjects under consideration suffer from the long time lasting experiments. They get tired, lose attention and can not adequately perform the experimental tasks anymore. As a consequence, data become less informative from the experimental design point of view. Furthermore, for some groups of probationers (infants or patients) long experiments may be too demanding.

Basically, we need less trials to shorten the time of the experiment. Hence, our attention is focused on methods, which extract useful information from EEG data more effectively than the conventional averaging does. This allows obtaining the desired accuracy of ERP characteristics using fewer trials and, hence, shorter experiment. We consider two methods that were developed to increase SNR in the subject averages as compared to the conventional averaging procedure: weighted averaging (Hoke et al., 1984) and ERPSUB (Ivannikov et al., 2007).

An important assumption underlying the averaging in electroencephalography research is the ergodicity of the noise. However, we should be realistic and

understand that this assumption is violated to some extent in practical applications. This leads us to a situation, when the variance of the noise is different across trials. It then turns out that SNR in the averaged responses can be boosted by weighting the trials inversely to the variance of the noise they contain. The formal derivation of this result was originally obtained in (Hoke et al., 1984) by minimizing the mean-square error criterion. In (Davila and Mobin, 1992) a similar technique has also been derived by maximizing SNR in the average using Rayleigh quotient and solving the generalized eigenvalue problem. Later, in (Łęski, 2002) robust version of weighted averaging was proposed and further developed into computationally more effective algorithm in (Łęski and Gacek, 2004). In this paper we obtain essentially same result as in (Hoke et al., 1984) by maximizing SNR criterion, but using different derivation procedure than that used in (Davila and Mobin, 1992) and show, thereby, that SNR criterion is equivalent to the mean-square error criterion in the sense of the weighted averaging problem solving.

ERPSUB method utilizes the problem specific assumptions for ERP/noise linear subspaces separation in multichannel EEG data and results in more effective denoising of ERPs comparing to the conventional averaging (Ivannikov et al., 2007). Method automatically solves the component classification problem for a priori known dimensionality of ERP subspace. Moreover, it contains also means for estimating the dimensionality of ERP subspace, if prior knowledge is absent, with the accuracy depending on the closeness of the data properties to the values provided by the ideal assumptions.

Since we are interested in decreasing the experimental time (minimizing number of trials necessary for reliable ERP identification), in this paper we concentrate on and compare the convergence rates of ERP estimates provided by selected methods (traditional averaging approach, weighted averaging, and ERPSUB), while the number of processed trials increases. Moreover, in order to give a comprehensive evaluation of the methods' performance, we also compare the noise reduction rates in ERP estimates for the same conditions.

The structure of the work is as follows. First, in Sect. 2, we describe the experimental data and formulate the research area. Then, in Sect. 3, the methods are discussed. Section 4 represents the experimental results. In Sect. 5, conclusions are drawn.

2 PRELIMINARIES

In this article we used EEG data that were introduced and studied in (Huttunen et al., 2007) and (Kalyakin et al., 2007). The same data were utilized also for the purposes of testing in (Ivannikov et al., 2007). The data collection experimental design was targeted to elicit mismatch negativity (MMN) component of auditory ERP. In fact, MMN has turned out to be especially useful for the investigation of the brain basis of human auditory cognition (Näätänen, 1992).

In the data collection experiments, the experimental paradigm proposed in (Pihko et al., 1995) was used. It is based on a sequence of standard stimuli consisting of continuously (uninterruptedly) alternated sounds of 600 Hz and 800 Hz, each lasting 100 ms. Two types of deviant stimuli are randomly presented in this sequence with the frequency of 600 Hz and duration of 30 ms or 50 ms. The measured trials contain 300 ms of recordings before the start of the deviant tone and 350 ms after the start of the deviant tone. Measurements were collected with the sampling rate 200 Hz, thus, giving 130 time points for each trial. There were 102 participants (or subjects) involved in the data collection experiment. Measurements were recorded using 12-electrodes scheme resulting in 350 trials collected for each of 102 subjects, each of the two deviants and each of the nine channels of EEG data (i.e., C3, C4, Cz, F3, F4, Fz, Pz, M1, M2) and the two channels of electrooculography (EOG) data (i.e., ER, EL). An additional nose electrode was used as a reference point.

We assume that each recorded trial $x_i^k(t)$ contains both the weighted sum of the time-locked brain responses $s^k(t)$ assumed to be deterministic through all trials and the weighted sum of the noise sources $n_i^k(t)$, such as spontaneous EEG and artifacts (Vigário, 1997; Jung et al., 2000). Noises $n_i^k(t)$ are assumed to be uncorrelated with each other and with $s^k(t)$. Then, without loss of generality we can assume that $x_i^k(t)$, $s^k(t)$, and $n_i^k(t)$ are zero mean variables, since data always can be centered. Hence, the simplest additive model to describe the phenomenon reads as

$$x_i^k(t) = s^k(t) + n_i^k(t), \qquad (1)$$

where $i = 1,...,N$, $t = 1,...,T$, and $k = 1,...,K$. Here N denotes the number of measured trials, T is the number of time points per trial, and K denotes the number of measured channels. The conventional averaging operation is performed for each channel separately and is described by formula:

$$\overline{x}_N^k(t) = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^k(t) = s^k(t) + \overline{n}_N^k(t), \qquad (2)$$

where $s^k(t)$ is the time-locked ERP constituent (signal of interest) and $\overline{n}^k_N(t)$ is the noise constituent in the average. The resulting average in (2) is assumed to have higher SNR than the single trial does that is confirmed by practical experience and theoretical computations (Näätänen, 1992; Furst and Blau, 1991).

3 METHOD DESCRIPTION

3.1 Weighted Averaging

The variance of the sum of N stochastic variables can be expressed through the formula

$$\sigma^2\left(\sum_{i=1}^{N} x_i(t)\right) = \sum_{i=1}^{N}\sigma^2(x_i(t)) + 2\sum_{i<j} Cov_{ij}, \quad (3)$$

where $Cov_{ij} = E[x_i(t)\,x_j(t)]$ denotes the covariance between the two zero mean stochastic variables or trials $x_i(t)$ and $x_j(t)$, and σ denotes the standard deviation. To simplify the following discussion we omit the channel index k throughout the paper assuming that all channels are treated in the similar way. Therefore, for the weighted sum/average of trials $\sum_{i=1}^{N} a_i x_i(t)$ and taking into account that the covariance of the two perfectly linearly correlated signals equals to the product of their standard deviations, we have

$$\sigma^2\left(\sum_{i=1}^{N} a_i x_i(t)\right) = \sum_{i=1}^{N} a_i^2\sigma_s^2 + \sum_{i=1}^{N} a_i^2\sigma_{n_i}^2 + 2\sigma_s^2\sum_{i=1}^{N-1} a_i \sum_{j=i+1}^{N} a_j, \quad (4)$$

where σ_s^2 denotes the variance of the signal and $\sigma_{n_i}^2$ is the variance of the noise in i-th trial. Then the portions of the total variance $\sigma_{\overline{s}}^2$ and $\sigma_{\overline{n}}^2$ that are contributed by the time-locked signals and noise sources, correspondingly, to the weighted sum (normal average in case $a_i = \frac{1}{N}, \forall i = 1,\ldots,N$) of N trials read as

$$\sigma_{\overline{s}}^2 = \sigma_s^2\left(\sum_{i=1}^{N} a_i\right)^2, \quad (5)$$

$$\sigma_{\overline{n}}^2 = \sum_{i=1}^{N} a_i^2\sigma_{n_i}^2. \quad (6)$$

We define SNR in the weighted sum of N trials as the variance of ERP constituent in this sum divided by the variance of the noise constituent:

$$SNR_N = \frac{\sigma_{\overline{s}}^2}{\sigma_{\overline{n}}^2} \quad (7)$$

and try to maximize its value in order to determine the optimal values of a_i's. For this purpose, taking the partial derivatives of $\sigma_{\overline{s}}^2$ and $\sigma_{\overline{n}}^2$ with respect to a_i, we have

$$\frac{\partial\,\sigma_{\overline{s}}^2}{\partial a_i} = 2\sigma_s^2\sum_{j=1}^{N} a_j, \forall 1 \le i \le N, \quad (8)$$

$$\frac{\partial\sigma_{\overline{n}}^2}{\partial a_i} = 2a_i\sigma_{n_i}^2, \forall 1 \le i \le N. \quad (9)$$

Therefore, the partial derivative of SNR with respect to a_i is given by

$$\frac{\partial SNR_N}{\partial a_i} = \frac{\frac{\partial\sigma_{\overline{s}}^2}{\partial a_i}\sigma_{\overline{n}}^2 - \sigma_{\overline{s}}^2\frac{\partial\sigma_{\overline{n}}^2}{\partial a_i}}{\left(\sigma_{\overline{n}}^2\right)^2}, \forall 1 \le i \le N. \quad (10)$$

Saddle points in the a_i's coordinate space can be found by equating the numerator of equation (10) to zero assuming $\sigma_{\overline{n}}^2 \neq 0$. Therefore, the problem can be expressed through a system of equations

$$\left(\sigma_s^2\sum_{j=1}^{N} a_j\right)\sigma_{\overline{n}}^2 - \sigma_{\overline{s}}^2(a_i\sigma_{n_i}^2) = 0, \quad \forall 1 \le i \le N. \quad (11)$$

Subtracting any two equations in this system, we obtain

$$a_i\sigma_{n_i}^2 = a_j\sigma_{n_j}^2, \forall 1 \le i,j \le N. \quad (12)$$

Plugging $a_j = \frac{\sigma_{n_i}^2}{\sigma_{n_j}^2}a_i$ back to the system of equations (11), we get a system of identical equations after some manipulations. Moreover, since the values of weighting coefficients a_i's were not fixed in this operation, they can be arbitrary within the constraint (12). This means, in turn, that

$$a_i\sigma_{n_i}^2 = a_j\sigma_{n_j}^2 = C, \forall 1 \le i,j \le N, \quad (13)$$

where C can be any constant. Hence, the solution has a form

$$a_i = \frac{C}{\sigma_{n_i}^2}, \forall 1 \le i \le N. \quad (14)$$

It is easy to check that this extremum point is the maximum by substituting $a_i = \frac{C}{\sigma_{n_i}^2} \pm \Delta, \forall 1 \le i \le N$ in (11), where $\Delta > 0$ is an infinitely small shift.

Assuming SNR in a single trial is very low (this follows from the magnitude level of the time-locked signal $\approx$ 3–5μV compared to the magnitude level of the trial itself $\approx$ 50–100μV), we can disregard the variance contributed by the time-locked signal to the trial and approximate

$$\sigma_{n_i}^2 \approx \sigma_{x_i}^2. \quad (15)$$

Thus, we can approximately compute the coefficients a_i's by arbitrarily fixing C constant first.

Note that in (Hoke et al., 1984) the minimization of the mean-square error leads to a single unique solution, whereas in our case the maximization of SNR yields an infinite set of solutions due to the arbitrary choice of C in (14). This result can be explained by the obvious reasoning that only the ratio between a_i's is emphasized by SNR criterion (the weighted sum can be multiplied by any number keeping SNR on a same level), whereas the solution based on minimizing mean-square error criterion is associated with the original level of ERP signal and with the highest SNR as well. Hence, in order to correct the level of ERP signal to original in the weighted average with weighting coefficients fixed as in (14), where C is arbitrary, we need to multiply $\sum_{i=1}^{N} a_i x_i(t)$ by a correction factor α that eliminates uncertainty introduced by arbitrariness of C. Apparently α depends on C and plays role of a constraint imposed on C and a_i's that specifies only single set of a_i's preserving the original level of ERP signal in the weighted average. From (5) α is obviously expressed through the formula

$$\alpha = \frac{1}{\sqrt{\sigma_{\bar{s}}^2/\sigma_s^2}} = \frac{1}{\sum_{i=1}^{N} a_i}. \tag{16}$$

After embedding the correction factor α into (14) the final solution for the weighting coefficients becomes

$$a_i = \sigma_{n_i}^{-2} / \sum_{j=1}^{N} \sigma_{n_j}^{-2}, \forall 1 \leq i \leq N \tag{17}$$

that coincides with the results from (Hoke et al., 1984). These values of the weighting coefficients are unique in the sense that they are connected to the original level of ERP signal and, thus, do not require multiplication by the correction factor α, which equals to 1 in this case.

3.2 ERPSUB

In the contemporary research EEG data is often considered in the scope of the linear instantaneous noiseless mixing model, which is also assumed in this paper:

$$X_i = A \cdot Y_i, \forall i = 1, \ldots, N, \tag{18}$$

where X_i is a matrix of size $K \times T$, which contains measurements from K channels and one trial of length T time points, Y_i is a matrix of size $K \times T$, which contains the realizations of K sources of length T time points, and A stands for the mixing matrix. It is assumed that every row in X_i has zero mean for all i, i.e. the data are centered. In addition we assume that the mixing matrix A does not change in time. Practically it means that for one subject during one experiment with the static conditions matrix A stays the same for all trials within the experiment. Therefore, we are allowed to form a data matrix by concatenating matrices X_i channel-wise:

$$X = A \cdot Y, \tag{19}$$

where $X = [X_1 \ X_2 \ \ldots \ X_N]$ is the matrix of concatenated measurements of size $K \times TN$ and $Y = [Y_1 \ Y_2 \ \ldots \ Y_N]$ is the matrix of concatenated realizations of the sources of the same size. Matrix equivalent of (2) can now be written as

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i = A \frac{1}{N} \sum_{i=1}^{N} Y_i = A\overline{Y}. \tag{20}$$

Furthermore, in the framework of the model (18) it is assumed that all K measurements in every multidimensional trial X_i are linearly independent and the number of sources does not exceed the number of channels. These assumptions are introduced to ensure that measurements form the basis for the linear space of the same dimension as sources do. This, in turn, guarantees the existence of the pure signal and noise subspaces in theory. Both assumptions are practically addressed by reasonable selection of K and T parameters. Moreover, we assume that subspaces of ERP signals and noise are statistically independent. The imposed assumptions, except the one concerning the linear independence of measurements, are rather strict and can not be completely justified in practical applications. However, they are necessary on the stage of the method development. In real situations one is instructed to reinterpret the results of the method according to the types and extent of the assumptions' violations.

The main idea of ERPSUB is to use the relevant information stored in data along all time, trial, and channel dimensions, while separating ERP/noise subspaces. In contrary, most of the Independent Component Analysis (ICA) methods also applied in EEG data processing to ERP/noise sources separation exploit the information kept along the time and channel dimensions only, whereas the trial dimension is ignored (Hyvärinen et al., 2001; Jung et al., 2000; Vigário, 1997). Traditional averaging is one-channel technique, and it exploits the information hidden in trial dimension only for ERP denoising. Weighted averaging is also one-channel procedure, but it utilizes the information taken from trial and time dimensions for the purposes of ERP denoising. ERPSUB exploits the fact that after the averaging the variance of data should decrease along the directions in the noise subspace, while the variance along the signal directions should stay on the original level in ideal conditions. This means that after whitening, which should make

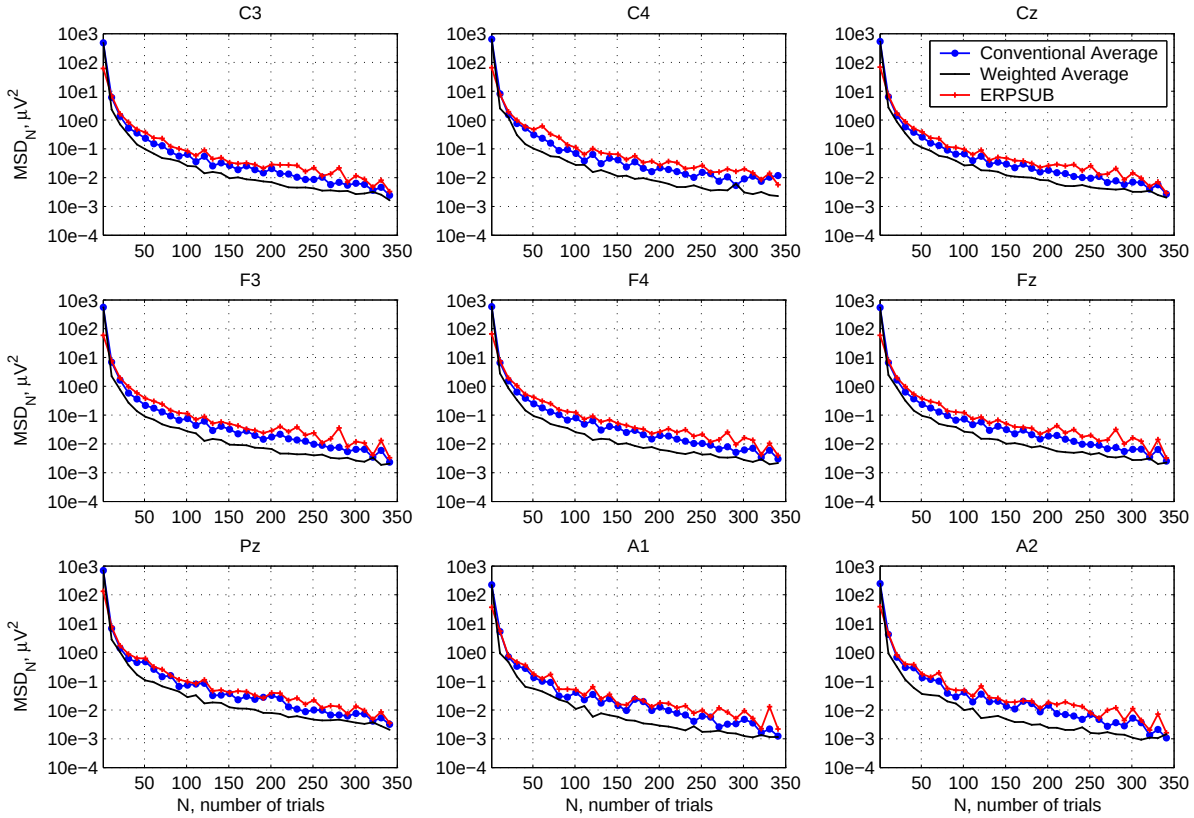


Figure 1: The averaged over 102 subjects MSD tracks provided by traditional averaging, weighted averaging and ERPSUB (nine EEG channels, 30ms deviant, logarithmic scale).

subspaces orthogonal and standardize the data to similar variances along all directions, and averaging ERP components should have the largest variances in contrary to the noise components, and, hence, subspaces can be extracted by standard linear Principal Component Analysis (PCA) algorithm (Hyvärinen et al., 2001; Oja, 1992). In practice, however, the variance of data most likely will reduce along all directions after the averaging, because subspaces are overlapped, and additive noise is always present, and, thus, pure signal/noise subspaces do not exist. In this case the results are interpreted in terms of SNR: higher SNR is obtained in data projected to the directions describing larger data variations after whitening and averaging. Thus, practically, we intend to separate the subspace of dimension N_{ERP} having maximal possible SNR from the subspace of dimension $K - N_{\mathrm{ERP}}$ with the minimal possible SNR. As one can see, ERPSUB is based on a sequence of linear transformations applied in a problem-specific manner to multidimensional EEG data and results in effective denoising of ERP signals (Ivannikov et al., 2007).

ERPSUB:

1. Whiten the centered concatenated data:
$$Z = D^{-1/2} W^{\mathrm{T}} X, \tag{21}$$
where matrices D and W are taken from the eigenvalue decomposition $\widehat{\Sigma} = WDW^{\mathrm{T}}$ of the estimated covariance matrix $\widehat{\Sigma} = XX^{\mathrm{T}}/(TN-1)$.
2. Average the whitened data:
$$\overline{Z} = D^{-1/2} W^{\mathrm{T}} \overline{X}. \tag{22}$$
3. Apply the standard linear PCA to the averaged whitened data
$$\overline{Y}'_{\mathrm{ERP}} = \Delta_{N_{\mathrm{ERP}}} \overline{W}^{\mathrm{T}} D^{-1/2} W^{\mathrm{T}} \overline{X}, \tag{23}$$
where matrix $\overline{W}^{\mathrm{T}}$ is obtained from the eigenvalue decomposition $\overline{Z}\overline{Z}^{\mathrm{T}}/(T-1) = \overline{W}D\overline{W}^{\mathrm{T}}$ and $\Delta_{N_{\mathrm{ERP}}}$ is the diagonal projection matrix having ones on N_{ERP} first diagonal elements corresponding to the components contributing energy to ERP (maximal SNR) subspace and zeros otherwise. Here, N_{ERP} is the amount of assumed ERP sources present in EEG measurements. In practice, when pure signal/noise subspaces do not exist, N_{ERP} has different meaning interpreted in terms of SNR. In this

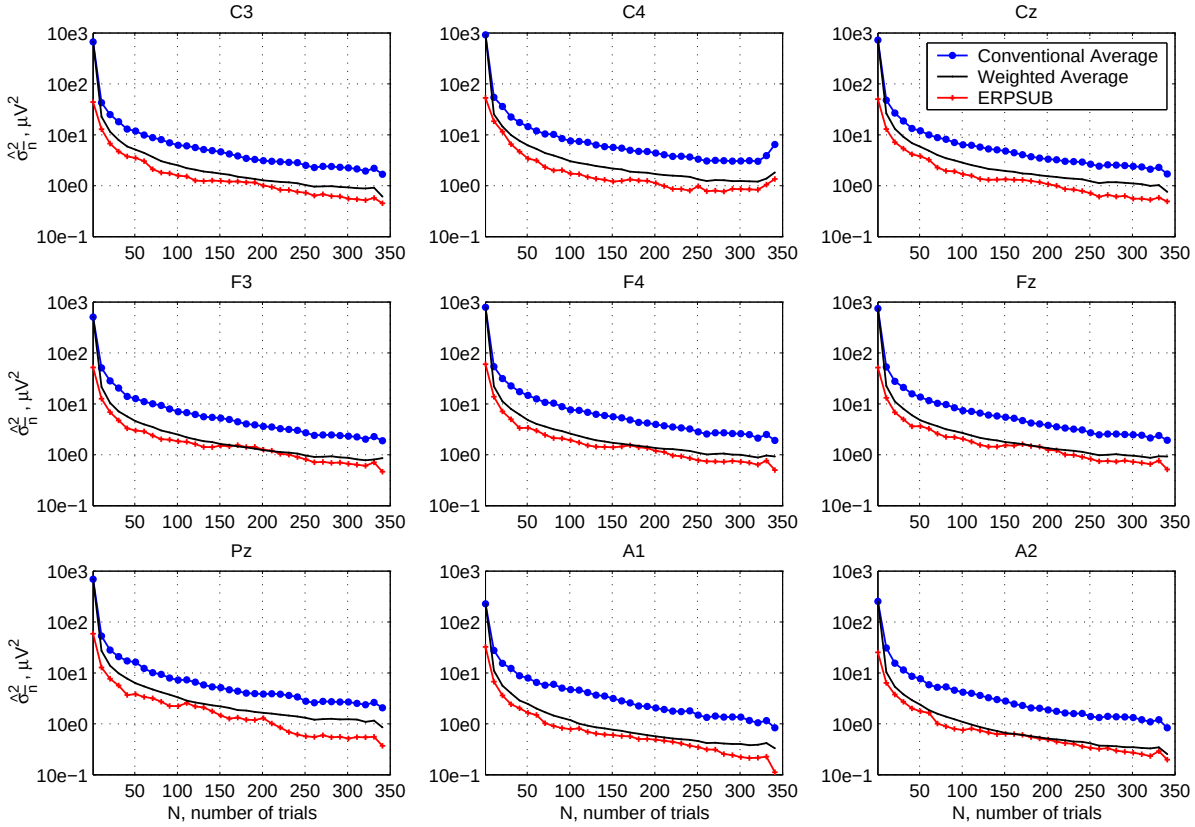


Figure 2: The averaged over 102 subjects tracks of the remaining noise variance in ERP estimates provided by traditional averaging, weighted averaging and ERPSUB (nine EEG channels, 30ms deviant, logarithmic scale).

case N_{ERP} is the amount of the components having largest SNR, which in our opinion describe ERP and noise variations in channels in proportions providing suitable SNR and tolerable ERP energy loss. Hence, $\overline{Y}'_{\mathrm{ERP}}$ is a matrix of the averaged components, where all components from noise (minimal SNR) subspace are zeroed. Note that ERP components have the largest corresponding eigenvalues and, thus, the component classification problem is solved automatically for fixed N_{ERP}. In addition, if the difference between eigenvalues corresponding to ERP and noise components is clearly observed, one can estimate N_{ERP} value providing optimal separation of the components into subspaces in the sense of SNR and ERP energy loss. Moreover, each K-dimensional trial X_i can be decomposed into the components using the same transformation as in (23):

$$Y'_{\mathrm{ERP}i} = \Delta_{N_{\mathrm{ERP}}} \overline{W}^{\mathrm{T}} D^{-1/2} W^{\mathrm{T}} X_i, \qquad (24)$$

4. The matrix $\overline{Y}'_{\mathrm{ERP}}$ containing only averaged components related to ERP subspace is then transformed back to the original data space (channels) to result in the subject average with the reduced noise:

$$\overline{X}_{\mathrm{ERP}} = W D^{1/2} \overline{W Y}'_{\mathrm{ERP}}. \qquad (25)$$

A similar relation applies also to a single trial denoising:

$$X_{\mathrm{ERP}i} = W D^{1/2} \overline{W} Y'_{\mathrm{ERP}i}. \qquad (26)$$

4 EXPERIMENTAL RESULTS

It was noticed during the simulations and is theoretically predictable that the weighted averaging method is highly sensitive to the trials having small portions of variance concentrated on short time intervals. Generally, such trials do not carry much of the information and are usually recorded at the saturation state of the amplifier, when parts of the trials are truncated resulting in peaks alternating with flat periods. Saturation state occurs, when signal exceeds the dynamical range of the amplifier. The weighting coefficients a_i's assigned for such trials are very large following the algorithm. As a consequence, when trials are weighted, peaks in truncated trials become very strong against a background of other trials' amplitude resulting in a

high frequency noise in the averaged signal. To annul the harmful consequences introduced by the truncated trials we performed the trial rejection procedure for our data before doing the computations. The successful upper limit of the trial's variance for the trial removal was $30\,\mu\mathrm{V}^2$, which finally rejected all truncated trials in our database.

Apparently, for our problem the converged ERP estimate (subject average) is indicated by only insignificant change introduced by the consequent trial. We measure the amount of change between the two subsequent ERP estimates in one channel for method LABEL by MSD score:

$$\mathrm{MSD}_N^{\mathrm{LABEL}} = \frac{1}{T}\sum_{t=1}^{T}(\overline{x}_N^{\mathrm{LABEL}}(t) - \overline{x}_{N-1}^{\mathrm{LABEL}}(t))^2,$$

where $\overline{x}_N^{\mathrm{LABEL}}(t)$ denotes ERP estimate obtained after application of a particular method LABEL involving N trials. Thus, for example, for ERPSUB $\overline{x}_N^{\mathrm{ERPSUB}}(t)$ equals to a row in the matrix of averaged filtered channels $\overline{X}_{\mathrm{ERP}}$ corresponding to the considered channel; for weighted averaging $\overline{x}_N^{\mathrm{WA}}(t) = \sum_{i=1}^{N}\widehat{a}_i x_i(t)$, where $\widehat{a}_i$'s are computed as in (17) substituting approximation from (15) for $\sigma_{n_i}^2$ for all $i = 1,\ldots,N$. To compare the convergence rates of ERP estimates provided by methods under consideration at increase of the number of processed trials, we computed averaged over 102 subjects MSD values for $N = 1,\ldots,350$ (MSD tracks) for each method (see Fig. 1). We did this for the nine EEG channels and for 30ms deviant only, where ERP appeared to be the strongest. The value of N_{ERP} parameter of ERPSUB method was set to 3, that is, a good choice of maximal SNR subspace dimension for our data, because signal loss is insignificant and noise reduction is sufficiently high resulting in essential SNR increase (Ivannikov et al., 2007). According to the obtained results, the weighted averaging procedure outperforms both the traditional averaging scheme and ERPSUB algorithm, because MSD provided by weighted averaging, in general, decreases faster than for other methods at increase of the number of processed trials. The superiority of the weighted averaging here is probably a consequence of the core idea underlying the method. Weighted averaging is designed in a way that trials are 'equalized' in the sense of the variance. This should make the convergence of the ERP estimate smoother and faster. Although application of ERPSUB should result in higher noise reduction rate than the conventional average provides (Ivannikov et al., 2007), ERPSUB has shown the lowest convergence rate of MSD to zero. Most likely this happens, because new-coming trial influences the denoising of all previous trials by changing the projection axes. Since the shapes of all filtered trials are affected, when new trial is added to processing, the difference between the two adjacent ERP estimates becomes more significant.

Therefore, in order to have a complete and fair comparative picture of the methods' performance, we also computed averaged over 102 subjects remaining noise variances in ERP estimates obtained under the same conditions as used in the first test (see Fig. 2). We used the following estimate of the noise variance in the averaged brain responses taken from (van de Velde, 2000):

$$\widehat{\sigma}_{\overline{n}}^{2\,\mathrm{LABEL}} = var\{\frac{1}{N}\sum_{i=1}^{N}(-1)^i x_i^{\mathrm{LABEL}}(t)\}, \qquad (27)$$

where $x_i^{\mathrm{LABEL}}(t)$ is the modified trial $x_i(t)$ obtained after application of method LABEL, and $\widehat{\sigma}_{\overline{n}}^{2\,\mathrm{LABEL}}$ is the estimate of the remaining noise variance in ERP estimate obtained after application of method LABEL involving N trials. For instance, $x_i^{\mathrm{WA}}(t) = \widehat{a}_i x_i(t)$, where $\widehat{a}_i$ is computed as in (17) replacing $\sigma_{n_i}^2$ with the approximation from (15) for all $i = 1,\ldots,N$; and $x_i^{\mathrm{ERPSUB}}(t)$ equals to a row in the matrix of filtered trial $X_{\mathrm{ERP}i}$ corresponding to the considered channel. In this test the performance order of the methods appeared to be different. ERPSUB has shown now the highest effectiveness in the sense of the noise reduction rate, since the remaining noise variance in ERP estimate provided by ERPSUB, in general, decreased faster than for other methods at increase of the number of processed trials. This outstanding performance can be explained here by the algorithmic nature of ERPSUB, which simultaneously operates through all time, trial, and channel dimensions that allows more efficient extraction of the information discriminating ERP and noise from data. The conventional averaging has shown the lowest noise reduction rate in ERP estimate following the results of the test.

5 CONCLUSIONS

In this article we compared the performance of the three methods used nowadays in EEG research for ERP denoising: conventional averaging, weighted averaging and ERPSUB. For this purpose we carried out two tests investigating the convergence and the noise reduction rates in ERP estimates provided by the selected methods at increase of the number of processed trials. The convergence rate of ERP estimate appeared to be the highest for the weighted averaging technique and the lowest for ERPSUB. However, ERPSUB has shown stronger noise reduction power than the traditional and weighted averaging methods have. The noise reduction rate in ERP estimate provided by the

traditional averaging was the poorest among tested methods.

The paper touches practical issues the neuropsychology researchers are faced with during EEG/ERP data processing and analyzing. Namely, it points out the bottlenecks of the traditional averaging technique used for the time-locked brain responses denoising. The roots of these bottlenecks are connected to the violation of the assumptions underlying the averaging in real applications and insufficiently powerful 'decoding' of the relevant information 'encrypted' in the data. The weighted averaging method addresses the bottlenecks, which arise due to the violation of the assumptions underlying traditional averaging. We have shown that this strategy for improving the performance and the reliability of the traditional averaging technique can be derived based on different criteria and, in particular, SNR and mean-square error as it has been shown in (Hoke et al., 1984). ERPSUB tries to eliminate the second type of the bottlenecks, which come from the disability of the traditional averaging to effectively extract from data the overall available information related to ERP and noise discrimination. ERPSUB 'decrypts' more effectively this information hidden in data, since it operates through all time, trial, and channel dimensions.

REFERENCES

Davila, C. E. and Mobin, M. S. (1992). Weighted averaging of evoked potentials. *IEEE Transactions on Biomedical Engineering*, 39(4):338–345.

Furst, M. and Blau, A. (1991). Optimal *a posteriori* time domain filter for average evoked potentials. *IEEE Transactions on Biomedical Engineering*, 38(9):827–833.

Hoke, M., Ross, B., Wickesberg, R., and Lütkenhöner, B. (1984). Weighted averaging - theory and application to electric response audiometry. *Electroenceph. Clin. Neurophysiol.*, 57:484–489.

Huttunen, T., Halonen, A., Kaartinen, J., and Lyytinen, H. (2007). Does mismatch negativity show differences in reading disabled children as compared to normal children and children with attention deficit? *Developmental Neuropsychology*, 31(3):453–470.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent component analysis*. John Wiley & Sons, Inc.

Ivannikov, A., Kärkkäinen, T., Ristaniemi, T., and Lyytinen, H. (2007). Extraction of ERP from EEG data. In *9th International Symposium on Signal Processing and its Applications*.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., and Sejnowski, T. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178.

Kalyakin, I., González, N., Joutsensalo, J., Huttunen, T., Kaartinen, J., and Lyytinen, H. (2007). Optimal digital filtering versus difference waves on the mismatch negativity in an uninterrupted sound paradigm. *Developmental Neuropsychology*, 31(3):429–452.

Łęski, J. (2002). Robust weighted averaging. *IEEE Transactions on Biomedical Engineering*, 49(8):796–804.

Łęski, J. and Gacek, A. (2004). Computationally effective algorithm for robust weighted averaging. *IEEE Transactions on Biomedical Engineering*, 51(7):1280–1284.

Luu, P., Tucker, D., and Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, 115(8):1821–35.

Makeig, S., Westerfield, M., Townsend, J., Jung, T.-P., Courchesne, E., and Sejnowski, T. (1999). Functionally independent components of the early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354:1135–44.

Näätänen, R. (1992). *Attention and brain function*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935.

Pihko, E., Leppäsaari, T., and Lyytinen, H. (1995). Brain reacts to occasional changes in duration of elements in continues sound. *NeuroReport*, 6:1215–18.

van de Velde, M. (2000). *Signal validation in electroencephalography research*. PhD thesis, Eindhoven University of Technology.

Vigário, R. (1997). Extraction of ocular artifacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103:395–404.

BIO-INSPIRED DATA AND SIGNALS CELLULAR SYSTEMS

André Stauffer, Daniel Mange and Joël Rossier
Logic Systems Laboratory, Ecole polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
andre.stauffer@epfl.ch

Keywords: Self-organization, configuration, cloning, cicatrization, regeneration.

Abstract: Living organisms are endowed with three structural principles: multicellular architecture, cellular division, and cellular differentiation. Implemented in digital according to these principles, our data and signals cellular systems present self-organizing mechanisms like configuration, cloning, cicatrization, and regeneration. These mechanisms are made of simple processes such as growth, load, branching, repair, reset, and kill. The data processed in the self-organizing mechanisms and the signals triggering their underlying processes constitute the core of this paper.

1 INTRODUCTION

Borrowing three structural principles (multicellular architecture, cellular division, and cellular differentiation) from living organisms, we have already shown how to grow cellular systems thanks to two algorithms: an algorithm for cellular differentiation, based on coordinate calculation, and an algorithm for cellular division (Mange et al., 2004). These cellular systems are endowed with self-organizing properties like configuration, cloning, cicatrization, and regeneration (Stauffer et al., 2005).

In a previous work (Stauffer et al., 2006), the configuration mechanisms (structural and functional growth), the cloning mechanisms (cellular and organismic growth), the cicatrization mechanism (cellular self-repair), and the regeneration mechanism (organismic self-repair) were already devised as the result of simple processes like growth, load, branching, repair, reset, and kill. The goal of this paper is to point out the data processed in these mechanisms and the signals triggering their underlying processes. Starting with a minimal system, a cell made up of six molecules, Section 2 will introduce digital simulations to describe the data and the signals involved in the self-organizing mechanisms and the corresponding processes. We define then a small organism made of three cells, the "SOS" acronym, as an application example for the simulation of our mechanisms and processes (Section 3). A brief conclusion (Section 4) summarizes our paper and opens new research avenues.

2 SELF-ORGANIZING MECHANISMS

2.1 Structural Configuration

The goal of the *structural configuration mechanism* is to define the boundaries of the cell as well as the living mode or spare mode of its constituting molecules. This mechanism is made up of a *structural growth process* followed by a *load process*. For a better understanding of these processes, we apply them to a minimal system, a cell made up of six molecules arranged as an array of two rows by three columns, the third column involving two spare molecules dedicated to self-repair.

The *growth process* starts when an external *growth signal* is applied to the lower left molecule of the cell (Fig. 1a) and this molecule selects the corresponding eastward data input (Fig. 1b). According to the *structural configuration data* or *structural genome*, each molecule of the cell generates then successively an internal *growth signal* and selects an input (Fig. 2), in order to create a data path among the molecules of the cell (Fig. 1b-g). When the connection path between the molecules closes, the lower left molecule delivers a *close signal* to the nearest left neighbor cell (Fig. 1h). The structural configuration data is now moving around the data path and ready to be transmitted to neighboring cells.

The *load process* is triggered by the *close signal* applied to the lower right molecule of the cell (Fig. 3a). A *load signal* propagates then westward and northward through the cell (Fig. 3b-d) and each of

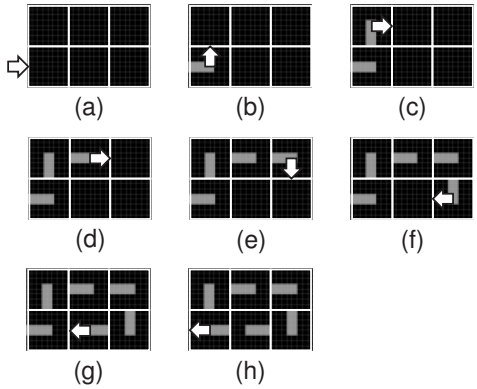


Figure 1: Structural growth process of a minimal system, a cell made up of six molecules. (a) External growth signal is applied to the lower left molecule. (b-g) Generation of internal growth signals to build the structural data path. (h) Closed path and close signal delivered to the nearest left neighbor cell.

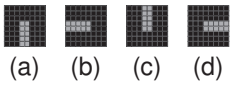


Figure 2: Data input selection. (a) Northward. (b) Eastward. (c) Southward. (d) Westward.

its molecules acquire a *molecular mode* (Fig. 4) and a *molecular type* (Fig. 5). We finally obtain an homogeneous tissue of molecules defining both the boundaries of the cell and the position of its *living mode* and *spare mode* molecules (Fig. 3e). This tissue is ready for being configured by the functional configuration data.

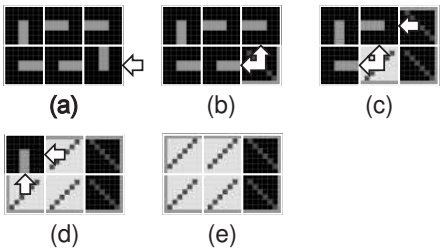


Figure 3: Load process. (a) External close signal applied to the lower right molecule by the nearest right neighbor cell. (b-e) Generation of internal load signals propagating westward and northward to store the molecular modes and types of the cell.

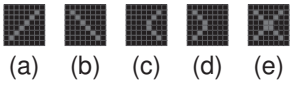


Figure 4: Molecular modes. (a) Living. (b) Spare. (c) Faulty. (d) Repair. (e) Dead.

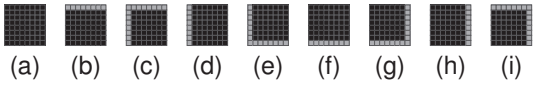


Figure 5: Molecular types. (a) Internal. (b) Top. (c) Top-left. (d) Left. (e) Bottom-left. (f) Bottom. (g) Bottom-right. (h) Right. (i) Top-right.

2.2 Functional Configuration

The goal of the *functional configuration mechanism* is to store in the homogeneous tissue, which already contains structural data (Fig. 3e), the functional data needed by the specifications of the current application. This mechanism is a *functional growth process*, performed only on the molecules in the *living mode* while the molecules in the *spare mode* are simply bypassed. It starts with an external *growth signal* applied to the lower left living molecule (Fig. 6a). According to the *functional configuration data* or *functional genome*, the living molecules then successively generate an internal *growth signal*, select an input, and create a path among the living molecules of the cell (Fig. 6b-f). The functional configuration data is now moving around the data path and ready to be transmitted to neighboring cells.

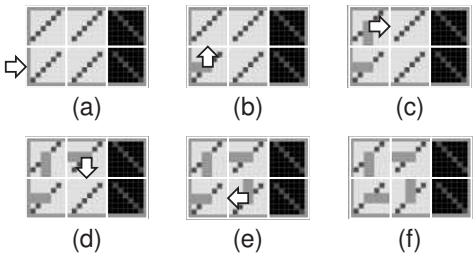


Figure 6: Functional configuration of the cell performed as a functional growth process applied to the living molecules. (a) External growth signal is applied to the lower left molecule. (b-e) Generation of internal growth signals in order to build the functional data path. (f) Closed functional data path.

2.3 Cloning

The *cloning mechanism* or *self-replication mechanism* is implemented at the cellular level in order to build a multicellular organism and at the organismic level in order to generate a population of organisms. This mechanism suppose that there exists a sufficient number of molecules in the array to contain at least one copy of the additional cell or of the additional organism. It corresponds to a *branching process* which

takes place when the structural and the functional configuration mechanisms deliver northward and eastward growth signals on the borders of the cell during the corresponding growth processes (Fig. 7).

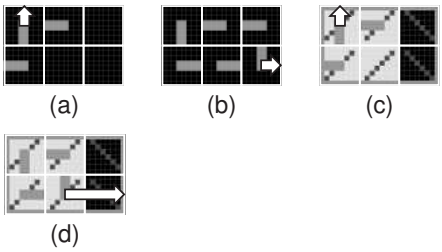


Figure 7: Generation of growth signals triggering the cloning mechanism. (a) Northward structural branching process. (b) Eastward structural branching process. (c) Northward functional branching process. (d) Eastward functional branching process.

2.4 Cicatrization

Fig. 6f, shows the normal behavior of a healthy minimal cell, i.e. a cell without any faulty molecule. A molecule is considered as faulty, or in the *faulty mode*, if some built-in self-test detects a lethal malfunction. Starting with the normal behavior of Fig. 6f, we suppose that two molecules will become suddenly faulty (Fig. 8a): (1) The lower left molecule, which is in the *living mode*. (2) The upper right molecule, which is in the *spare mode*. While there is no change for the upper right molecule, which is just no more able to play the role of a spare molecule, the lower left one triggers a *cicatrization mechanism*. This mechanism is made up of a *repair process* involving eastward propagating *repair signals* (Fig. 8b-c) followed by a *reset process* performed with westward and northward propagating internal *reset signals* (Fig. 8d-g). This tissue, comprising now two molecules in the *faulty mode* and two molecules in the *repair mode*, is ready for being reconfigured by the functional configuration data. This implies a *functional growth process* bypassing the faulty molecules (Fig. 9).

2.5 Regeneration

Our minimal system comprises a single spare molecule per row and tolerates therefore only one faulty molecule in each row. A second faulty molecule in the same row will cause the death of the whole cell, and the start of a *regeneration mechanism*. Fig. 10 illustrates the *repair process* and *kill process* involved in this mechanism. Starting with the normal behavior of the cicatrized cell (Fig. 9f), a new

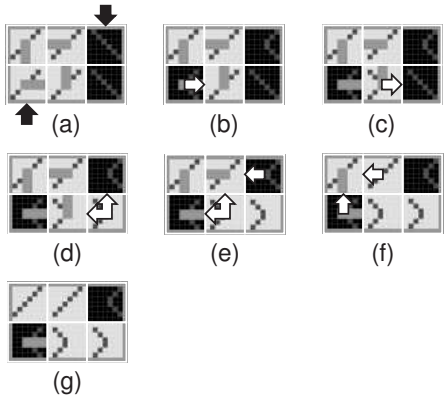


Figure 8: Cicatrization mechanism performed as a repair process followed by a reset process. (a) Living and spare molecules becoming faulty. (b-c) Generation of repair signals propagating eastward. (d-f) Generation of internal reset signals propagating westward and northward. (g) Cell, comprising two faulty and two repair molecules, ready for functional reconfiguration.

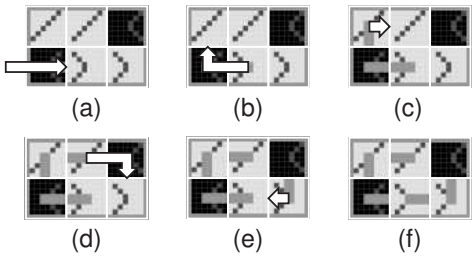


Figure 9: Functional reconfiguration of the living and repair molecules. (a) External growth signal bypassing the lower left faulty molecule. (b-e) Generation of internal growth signals to build a functional data path bypassing the faulty molecules. (f) Closed functional data path within the living and repair molecules.

molecule, the upper middle one, becomes faulty. In a first step, the new faulty molecule sends a *repair signal* eastward, in order to look for a spare molecule, able to replace it (Fig. 10b). In a second step, the supposed spare molecule, which is in fact a faulty one, enters the lethal *dead mode* and triggers *kill signals* which propagate northward, westward and southward (Fig. 10c-f). Finally in Fig. 10g, all the molecules of the array are dead as well as our minimal system.

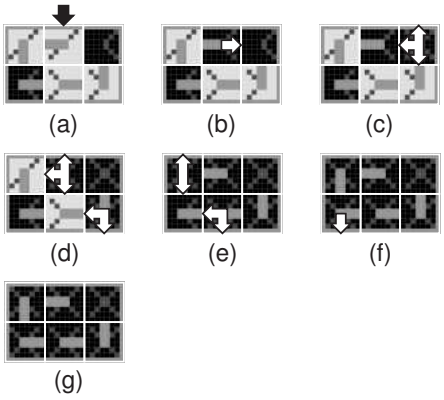


Figure 10: Regeneration mechanism performed as a repair process followed by a kill process. (a) Living molecule becoming faulty. (b) Eastward repair signal. (c-f) Generation of internal and external kill signals propagating northward, westward and southward. (g) Cell made up six dead molecules.

3 SOS ACRONYM APPLICATION

3.1 Structural Configuration, Functional Configuration and Cloning

Even if our final goal is the self-organization of complex bio-inspired data and signals cellular systems, we will use an extremely simplified application example, the display of the "SOS" acronym, in order to illustrate its basic mechanisms. The system that displays the acronym can be considered as a one-dimensional artificial organism composed of three cells (Fig. 11). Each cell is identified by a X coordinate, ranging from 1 to 3. For coordinate values $X = 1$ and $X = 3$, the cell implements the S character, for $X = 2$, it implements the O character. Such a cell, capable of displaying either the S or the O character, is a *totipotent cell* comprising $4 \times 6 = 24$ molecules.

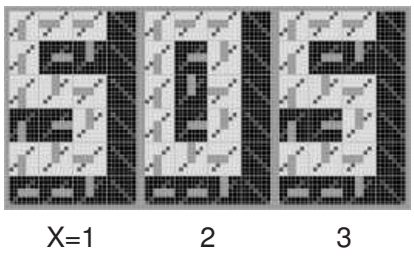


Figure 11: One-dimensional organism composed of three cells resulting from the structural configuration, functional configuration and cloning mechanisms applied to a totipotent cell.

In order to build the multicellular organism of Fig. 11, the structural configuration mechanism, the functional configuration mechanism, and the cloning mechanism are applied at the cellular level. Starting with the structural and functional configuration data of the totipotent cell, these mechanisms generate successively the three cells $X = 1$ to $X = 3$ of the organism "SOS".

3.2 Cicatrization and Functional Reconfiguration

The cicatrization mechanism (or cellular self-repair) results from the introduction in each cell of one column of spare molecules (Fig. 11), defined by the structural configuration of the totipotent cell, and the automatic detection of faulty molecules. Thanks to this mechanism, each of the two faulty molecules of the middle cell (Fig. 12) is deactivated, isolated from the network, and replaced by the nearest right molecule, which will itself be replaced by the nearest right molecule, and so on until a spare molecule is reached. The functional reconfiguration mechanism takes then place in order to regenerate the O character of the organism "SOS". As shown in Fig. 12, the regenerated character presents some graphical distortion.

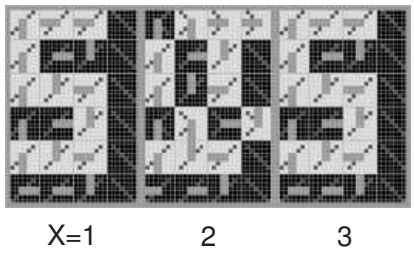


Figure 12: Graphical distortion resulting from the cicatrization and reconfiguration mechanisms applied to the middle cell of the organism.

3.3 Regeneration

The totipotent cell of the organism "SOS" having only one spare column allows only one faulty molecule per row. When a second one is detected, the regeneration mechanism (or organismic self-repair) takes place and the entire column of all cells to which the faulty cell belongs is considered faulty and is deactivated (column $X = 2$ in Fig. 13; in this simple example, the column of cells is reduced to a single cell). All the functions (X coordinate and configuration) of the cells to the right of the column $X = 1$ are shifted by one column to the right. Obviously, this process requires as

many spare cells to the right of the array as there are faulty cells to repair. As shown in Fig. 13, the reparation of one faulty cell needs one spare cell to the right and leaves a scar in the organism "SOS".

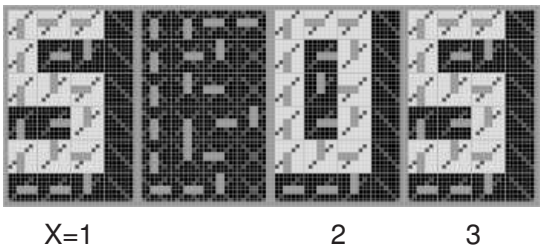


Figure 13: Scar resulting from the regeneration mechanism applied to the organism.

4 CONCLUSIONS

The self-organizing mechanisms are made of simple processes like growth, load, branching, repair, reset, and kill. They allow the data and signals cellular systems to possess three bio-inspired properties: (1) Cloning or self-replication at cellular and organismic levels. (2) Cicatrization or self-repair at the cellular level. (3) Regeneration or self-repair at the organismic level.

Starting with a very simple system, a cell made of six molecules, we realized digital simulations in order to describe the data and signals involved in the self-organizing mechanisms. The "SOS" acronym, an organism made of three cells, was introduced as an application example for the simulation of our mechanisms and processes.

The functional configuration mechanism presented here will be implemented in the *ubichip* (Upegui et al., 2007), a programmable circuit that draws inspiration from the multi-cellular structure of complex biological organisms.

REFERENCES

Mange, D., Stauffer, A., Petraglio, E., and Tempesti, G. (2004). Self-replicating loop with universal construction. *Physica D*, 191(1-2):178–192.

Stauffer, A., Mange, D., and Tempesti, G. (2005). Embryonic machines that grow, self-replicate and self-repair. In J. Lohn, e. a., editor, *Proceedings of the 2005 NASA/DoD Conference on Evolvable Hardware (EH'05)*, pages 290–293. IEEE Computer Society, Los Alamitos, CA.

Stauffer, A., Mange, D., and Tempesti, G. (2006). Bio-inspired computing machines with self-repair mechanisms. In Ijspert, A., Masuzawa, T., and Kusumoto, S., editors, *Biologically Inspired Approaches to Advanced Information Technology. Proceedings of The Second International Workshop Bio-ADIT 2006*, Lecture Notes in Computer Science. Springer-Verlag, Heidelberg.

Upegui, A., Thomas, Y., Sanchez, E., Perez-Uribe, A., Moreno, J.-M., and Madredas, J. (2007). The Perplexus bio-inspired reconfigurable circuit. In J. Lohn, e. a., editor, *Proceedings of the NASA/ESA Conference on Adaptative Hardware and Systems (AHS-2007)*, pages 600–605. IEEE Computer Society, Los Alamitos, CA.

ANALYSIS OF DIFFERENCES BETWEEN SPECT IMAGES OF THE LEFT AND RIGHT CEREBRAL HEMISPHERES IN PATIENTS WITH EPILEPTIC SYMPTOMS

Elżbieta Olejarczyk and Małgorzata Przytulska
Institute of Biocybernetics and Biomedical Engineering PAS, ul. Ks.Trojdena 4, 02-109 Warszawa, Poland
eolejarczyk@ibib.waw.pl, gosia@ibib.waw.pl

Keywords: SPECT, epilepsy, fractal dimension, entropy.

Abstract: The aim of his work was examination of asymmetries in activity of the left and right cerebral hemispheres as well as localization and contouring of the regions of reduced or increased activity on the basis of single photon emission computer tomography (SPECT) images. The mean and standard deviation of normalized intensities inside the contoured areas of images, entropy based on intensity histograms and Chen's fractal dimension were calculated.

1 INTRODUCTION

The aim of his work was examination of asymmetries in activity of the left and right cerebral hemispheres as well as localization and contouring of the regions of reduced or increased activity on the basis of single photon emission computer tomography (SPECT) images (Prószyński, 1989). Advantage of this technique is possibility of brain activity map acquisition at the time of radiotracer injection during seizures though the image registration is done one hour after seizure. SPECT imaging method allows better spatial localization of seizure source than the analysis of EEG signal. Simultaneous EEG signal registration allows to qualify exactly the moment of seizure onset when radiotracer injection could be done to register an unequivocal image. The mean and standard deviation of normalized intensities inside the contoured areas of images were calculated. Methods like entropy based on intensity histograms and Chen's fractal dimension were also applied.

2 MATERIALS

The scintigraphic examinations of cerebral perfusion in 6 patients were performed in the Department of Nuclear Medicine of the Medical Academy of Warsaw.

From each patient after delivering them the HMPAO Tc99m isotope in interictal phase several transverse cerebral images have been acquired. In the below-shown series of images they have been ordered from the basis to the top of the examined brain; left side of an image corresponds to the right side of the brain and vice versa.

An increased/reduced cerebral perfusion corresponds to a higher/lower isotope density and is manifested by an increased/ reduced image luminance registered in an 8-bits scale and normalized to the maximum (256 steps) luminance level. Images of 128x128 pixels size were registered. In table 1 several examples of medical description of the corresponding cases are given.

3 METHODS

In order to evaluate the effectiveness of various methods the comparative analysis of the images of the left and right cerebral hemispheres was performed by using three independent methods:

1. comparison of the mean and standard deviation values,
2. comparison of estimated entropies,
3. comparison of fractal dimensions.

Table 1: Kind of perfusion and localization of the brain region by medical assessment.

Patent name (slices number)	Kind of perfusion	Localization on the image	Brain region
CHM (16)	Reduced	Right upper	Left frontal lobe
KOS (15)	Disabled	Right	Left temporal lobe
SIE (15)	Reduced	Right	Left temporal lobe
SZY (14)	Disabled	Left upper	Right frontal temporal lobe
TWO (15)	Disabled	Left upper	Right frontal temporal lobe
ZIE (11)	Increased	Left	Right temporal lobe (numerous movement artifacts)

3.1 The Mean and Standard Deviation of Normalized Intensities

The images were processed and analyzed using standard Image Pro Plus (Media Cybernetics) and Microsoft Excel software packages (Russ, 1995). Each image was geometrically divided into the left and right parts. For a direct visual assessment of monochromatic images they also were visualized in pseudocolors. Then the left and right cerebral hemispheres were automatically contoured and the mean and standard deviation of normalized luminances inside the contoured areas of images were calculated. At the next step the surrounding background, outside a mask selecting the object of interest, from the images was reduced to the 0 level.

3.2 Entropy based on Intensity Histograms

The Shannon's entropy (Shannon, 1948) of a probability distribution of image intensities is defined as:

$$S = -\sum^{N}_{i=1} p_i \log p_i \quad , \quad \sum^{N}_{i=1} p_i = 1 \qquad (1)$$

where

p_i , i=1,…N - probability of i-th intensity level.

Entropy based on intensity histogram (Kuczyński and Mikołajczyk, 2003) can be estimated as:

$$p_i = g_i / g_{total} \qquad (2)$$

where g_i - number of pixels with intensity i;
g_{total} - total number of pixels;
N - number of image gray levels.

Entropy is a measure of information. The bigger are changes of pixel intensities the bigger is the entropy. In this method only a total histograms are used to calculate entropy therefore the spatial information is lost.

3.3 Chen's Fractal Dimension

For image matrices with dimension N x N a multi-scale vector of difference intensity MSID = [ri(s1),ri(s2),… ri(sk)], where ri(sk) – mean intensity of all pairs of pixels at the distance sk, was defined (Chen, 1989).

If I(x,y) is a measure of intensity (gray level at point with (x,y) coordinates), then:

$$ri(sk) = \sum^{N-1}_{x1=0} \sum^{N-1}_{y1=0} \sum^{N-1}_{x2=0} \sum^{N-1}_{y2=0} (|I(x2,y2) - I(x1,y1)|) \ / \ \text{number of pixel pairs for sk scale} \qquad (3)$$

There are the following relations for coordinates x1, y1, x2, y2:

$$sk = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \qquad (4)$$

$$|\Delta I| = |I(x2,y2) - I(x1,y1)| \sim |\Delta x|^H \qquad (5)$$

where
H – Hurst's exponent (fractal dimension Df = 3-H);
Δx – the distance between points with coordinates (x2,y2) and (x1,y1).

The logarithms of both sides were calculated:

$$\log |\Delta I| \sim H \cdot \log |\Delta x| \qquad (6)$$

4 RESULTS

4.1 The Mean and Standard Deviation of Normalized Intensities

The comparative analysis was performed in 6 patients for which mean values and standard deviations of luminance on the left and right cerebral hemispheres were measured. In order to make the results independent on the mean brightness level data were normalized by calculation of the ratio of

the difference to the sum of mean brightness in the hemispheres. The regions of reduced/increased perfusion were localized using the above-described image segmentation method. The results of calculations (for two patients mentioned above) are shown in Fig. 1. The horizontal axes indicate the numbers of consecutive slices. The normalized values of mean brightness differences for 6 patients are shown in Fig. 2.

4.2 Entropy and Fractal Dimension

Entropy and Chen's fractal dimension were calculated for all quarters of brain (upper-left, upper-right, down-left, down-right) in 8 ranges of pixel intensity: 1-32, 33-64, 65-96, 97-128, 129-160, 161-192, 193-224, 225-256. Each of four regions of brain contains 63 x 63 pixels. The ratio of the difference

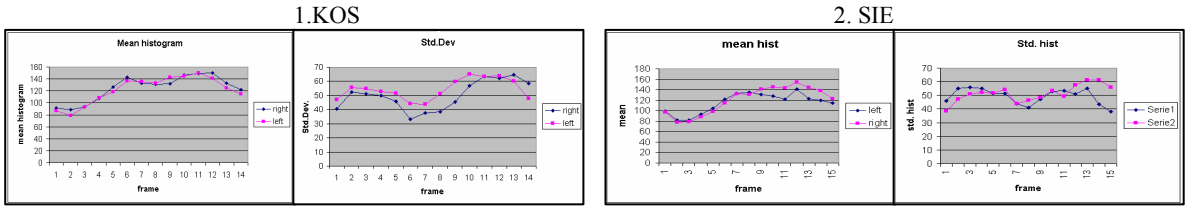


Figure 1: Mean values and standard deviations of luminance for the left and right cerebral hemispheres, patients KOS, SIE.

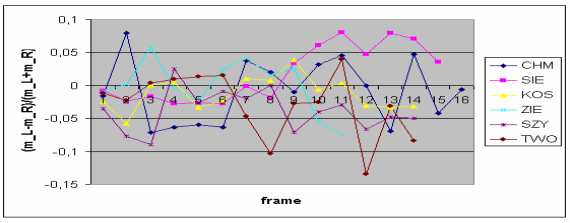


Figure 2: Normalized values of mean brightness differences for 6 patients.

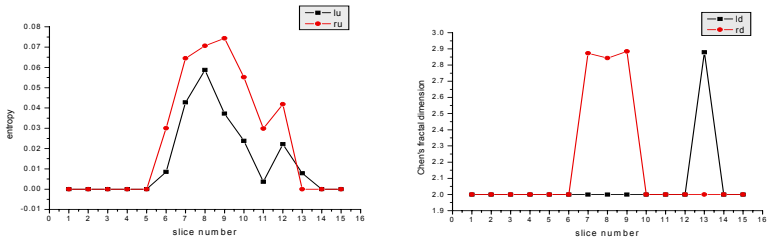


Figure 3: Entropy and Chen's fractal dimension graphs for the higher level of pixels intensity (225-256) for patient KOS in regions of brain in which differences of these measures between both hemispheres are bigger than 10%.

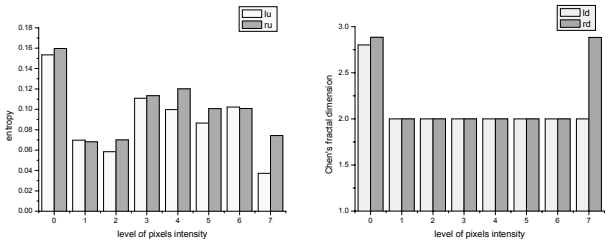


Figure 4: Histogram of entropy and Chen's fractal dimension for all levels of pixels intensity for patient KOS in regions of brain in which differences of these measures between both hemispheres are bigger than 10%.

to the sum of entropy and Chen's fractal dimension for left upper/down respect of right upper/down quarter of brain was calculated. Entropy results confirmed the medical observations (Table 2). Graphs of entropy and Chen's fractal dimension for the higher level of pixel intensity (225-256) in regions of brain for which the differences of these measures between both hemispheres are bigger than 10% are showed on Fig. 3. Histograms of entropy and fractal dimension (Fig.4) show significant differences for the higher level of pixel intensity.

Table 2: Number of slices for which the rate of the difference to the sum of S or Df in the left and right hemispheres in range with the biggest intensity of pixels (from 193 to 256) has values bigger than 0.1 for four regions of brain (UL-upper left, UR-upper right, DL-down left, DR-down right).

Pacjent name	Number of slices for entropy UL-UR-DL-DR	Slices number for fractal dimension UL-UR-DL-DR
CHM	1-7-7-0	1-1-0-4
KOS	3-7-8-0	6-1-1-3
SIE	0-9-3-5	1-0-0-0
SZY	10-0-9-0	0-0-0-0
TWO	3-3-2-0	1-0-0-5
ZIE	1-7-4-2	1-1-0-0

5 CONCLUSIONS

The above presented methods of cerebral SPECT images analysis based on simple image processing methods and calculation of basic statistical parameters are effective tools for a preliminary assessment of cerebral perfusion in diagnosis of epileptic and/or cerebral ischemic patients. It was found that for reduced perfusion entropy increases and Chen's fractal dimension decreases. Entropy based on the intensity histograms permits on automatic perfusion asymmetry evaluation between left and right brain hemisphere taking into account only the bigger intensities of pixels (in the range from 193 to 256). Entropy is a better measure to estimate the global intensity however without information about spatial distribution. For identification of epileptic seizure localization (concentration of high intensity pixels) Chen's fractal dimension seems to be the better measure.In further work calculations for more patients and for group of healthy volunteers should be done. Chen's fractal dimension could be calculated for less-dimensional matrices (8 x 8) in sliding window to construct map of fractal dimension of the whole brain. It will allow to estimate better the utility of this method to localize the epileptic seizure and to compare different regions of interest (ROIs).

ACKNOWLEDGEMENTS

We acknowledge thanks to prof. Leszek Królicki and dr Adam Bajera from the Department of Nuclear Medicine of the Medical Academy of Warsaw for providing databases – the SPECT images registered for epileptic patients.

This work was supported by Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences under Grant St/18/07 and ST/21/07.

REFERENCES

Chen C., Daponte J., Fox M., 1989. Fractal feature analysis and classification in medical imaging. *IEEE Trans. Med.. Imaging, 8, 133-142.*

Kuczyński K., Mikołajczyk P., 2003. Dopasowanie i segmentacja obrazów medycznych w oparciu o elementy teorii informacji (in polish). In *XIII Krajowa Konferencja Naukowa Biocybernetyka i Inżynieria Biomedyczna*, *Gdańsk, pp. 806-811.*

Oczeretko E.S., 2006. *Wymiar fraktalny w analizie sygnałów i obrazów biomedycznych (in polish)*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok.

Shannon C.E., 1948. A Mathematical Theory of communication. *The Bell System Technical Journal*, *27, 379-423, 623-659.*

Prószyński B., 2006. *Radiologia obrazowa RTG, TK, USG, MR i radioizotopy (in polish),* PZWL, Warszawa.

Russ J.C., 1995. *Image processing handbook*. 2nd edition. CRC Press, Boca Raton, Ann Arbor, London.

MOTION ESTIMATION IN MEDICAL IMAGE SEQUENCES USING INVERSE POLYNOMIAL INTERPOLATION

Saleh Al-Takrouri
School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW, Australia and National ICT Australia Ltd, Sydney, NSW, Australia
saleh@student.unsw.edu.au

Andrey V. Savkin
School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW, Australia
a.savkin@unsw.edu.au

Keywords: Motion estimation, polynomial interpolation, inverse interpolation, medical image sequences.

Abstract: We propose a new method for motion estimation between two successive frames in medical image sequences and videos. The method is based on inverse polynomial interpolation.

1 INTRODUCTION

The applications of motion estimation have been increasingly gaining interest in the field of medical imaging. (Hemmendorff, 2001) proposed a framework for motion estimation of 2D X-ray angiography images and 3D MRI mammograms. Deformable models were used by (Kurabayashi et al., 2005) to estimate the motion in time-series chest MR images. (Auvray et al., 2006) applied motion estimation to transparent X-ray image sequences.

Motion estimation is a key step in video coding and compression, which is an important tool to achieve bandwidth reduction when transmitting medical image sequences and videos. In addition, remote and robot-assisted surgeries and medical diagnostic tools can benefit from motion estimation in analyzing and interpreting the motions of body parts.

2 PROBLEM STATEMENT

Consider the pair of images $I_1(r,c)$ and $I_2(r,c)$, both of size $R \times C$, where the spacial arguments r and c refer to the pixel at the rth row and cth column. Here we assume that the two images are successive frames in a medical video or image sequence with spacial differences between the two images but no change in intensity. The pixels $I_2(r,c)$ of the destination image can be generated by shifting corresponding pixels in the source image in the 2D space. Let $\tau_1(r,c)$ and $\tau_2(r,c)$ be the horizontal and vertical shifts respectively, then we can write

$$I_2(r,c) = I_1(r+\tau_2(r,c), c+\tau_1(r,c)) \qquad (1)$$

Sub-pixel shifts are approximated by 2D polynomial interpolations within square neighborhoods of the source image I_1. The advantage of this choice is the separability and simplicity of implementation that allows an approximation of (1) to be written in an easily manipulated form.

Now assume that $|\tau_1(r,c)| \leq p$ and $|\tau_2(r,c)| \leq p$. Then the neighborhood in consideration would be of size $(2p+1) \times (2p+1)$ and the interpolation polynomial is of order $2p$. Define a vector function

$$\mathbf{u}(\tau_i(r,c)) = \begin{bmatrix} \tau_i^{2p}(r,c) \\ \tau_i^{2p-1}(r,c) \\ \vdots \\ 1 \end{bmatrix} \qquad (2)$$

The polynomial approximation of (1) can be written in the form

$$I_2(r,c) = \mathbf{u}^T(\tau_1(r,c))\mathbf{A}(r,c)\mathbf{u}(\tau_2(r,c)) \qquad (3)$$

where $\mathbf{A}(r,c)$ is a $(2p+1) \times (2p+1)$ matrix . For simplicity, the spacial arguments (r,c) are dropped from this point and assumed implicitly

$$I_2 = \mathbf{u}^T(\tau_1)\mathbf{A}\mathbf{u}(\tau_2) \qquad (4)$$

With τ_1 and τ_2 are the unknowns in equation (4), our goal is to solve the inverse polynomial interpolation problem represented by (4), which would also solve the motion estimation problem described above.

Many motion estimation methods use multiscale or hierarchial levels in order to process large motions, the proposed method can handle the size of motions that typically exist between two successive frames and therefor we are not using any multiscale pyramids.

3 POLYNOMIAL INTERPOLATION

For a pixel that is assumed to move a maximum of p pixels to the right or the left in a 1D source signal, the neighborhood considered $\mathbf{Y}$ is of length $2p+1$ and centered at the element $y(0)$. Using polynomial interpolation, $y(x)$ representing a shift from the center by a value x where $|x| \leq p$ can be approximated by using a polynomial of order $2p$

$$y(x) = c_{2p}x^{2p} + c_{2p-1}x^{2p-1} + \cdots + c_2x^2 + c_1x + c_0 \tag{5}$$

The coefficients $c_{2p} \cdots c_0$ are found by solving a system of $2p+1$ linear equations of the form

$$\mathbf{XC} = \mathbf{Y}^T \tag{6}$$

where

$$\mathbf{X} = \begin{bmatrix} (-p)^{2p} & (-p)^{2p-1} & \cdots & -p & 1 \\ (-p+1)^{2p} & (-p+1)^{2p-1} & \cdots & -p+1 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ (p-1)^{2p} & (p-1)^{2p-1} & \cdots & p-1 & 1 \\ p^{2p} & p^{2p-1} & \cdots & p & 1 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} c_{2p} \\ c_{2p-1} \\ \vdots \\ c_1 \\ c_0 \end{bmatrix}, \qquad \mathbf{Y}^T = \begin{bmatrix} y(-p) \\ \vdots \\ y(0) \\ \vdots \\ y(p) \end{bmatrix} \tag{7}$$

and the solution to the linear system is given by

$$\mathbf{C} = \mathbf{QY}^T, \qquad \mathbf{Q} = \mathbf{X}^{-1} \tag{8}$$

The matrix $\mathbf{X}$ in (7) is a special form of the Vandermonde matrix. Its inverse can be found using an explicit *LU* factorization discussed in the paper by (Olver, 2006).

Denote the ith row of the matrix $\mathbf{Q}$ in (8) as $\mathbf{q}_i$. The process of 1D polynomial interpolation can be expressed as

$$y(x) = \mathbf{Y} \sum_{i=1}^{2p+1} \mathbf{q}_i^T x^i \tag{9}$$

The 1D polynomial interpolation in (9) can be extended to the 2D case. When a pixel in a 2D neighborhood is assumed to move a maximum of p pixels along any dimension, the neighborhood in consideration is of size $(2p+1) \times (2p+1)$ and centered at the pixel $n(0,0)$.

Recall that $\mathbf{q}_i$ is the ith row of the matrix $\mathbf{Q}$ in (8). We use the fact that the 2D polynomial interpolation is separable to build the matrix $\mathbf{A}$, with each element on the ith row and jth column is given by

$$a_{(i,j)} = \sum_{m=-p}^{p} \sum_{n=-p}^{p} \mathbf{N}(m,n)\mathbf{q}_j(m)\mathbf{q}_i(n) \tag{10}$$

The process of 2D polynomial interpolation can be expressed now as

$$I_2 = \sum_{i=1}^{2p+1} \sum_{j=1}^{2p+1} a_{(i,j)} \tau_1^{2p+1-i} \tau_2^{2p+1-j} \tag{11}$$

with the matrix form of (11) is as given by (4).

4 SOLUTION OF INVERSE INTERPOLATION

4.1 The Linear Approximation

First we start by finding a linear approximation of (4) around some values $\bar{\tau}_1, \bar{\tau}_2$ (to be defined later). The first order approximation using Taylor series is easily computed since the differentiation of (4) with respect to ether τ_1 or τ_2 is trivial.

$$\begin{aligned} I_2 \approx \quad & \mathbf{u}^T(\bar{\tau}_1)\mathbf{A}\mathbf{u}(\bar{\tau}_2) + \dot{\mathbf{u}}^T(\bar{\tau}_1)\mathbf{A}\mathbf{u}(\bar{\tau}_2)\,[\tau_1 - \bar{\tau}_1] \\ & + \mathbf{u}^T(\bar{\tau}_1)\mathbf{A}\dot{\mathbf{u}}(\bar{\tau}_2)\,[\tau_2 - \bar{\tau}_2] \end{aligned} \tag{12}$$

Equation (12) is written in a form of a linear equation

$$\bar{I}(\bar{\boldsymbol{\tau}}) = \mathbf{H}(\bar{\boldsymbol{\tau}})\boldsymbol{\tau} \tag{13}$$

where

$$\begin{aligned} \bar{I}(\bar{\boldsymbol{\tau}}) &= I_2 - \mathbf{u}^T(\bar{\tau}_1)\mathbf{A}\mathbf{u}(\bar{\tau}_2) + \mathbf{H}(\bar{\boldsymbol{\tau}})\bar{\boldsymbol{\tau}} \\ \mathbf{H}(\bar{\boldsymbol{\tau}}) &= \begin{bmatrix} \dot{\mathbf{u}}^T(\bar{\tau}_1)\mathbf{A}\mathbf{u}(\bar{\tau}_2) & \mathbf{u}^T(\bar{\tau}_1)\mathbf{A}\dot{\mathbf{u}}(\bar{\tau}_2) \end{bmatrix} \\ \bar{\boldsymbol{\tau}} &= \begin{bmatrix} \bar{\tau}_1 \\ \bar{\tau}_2 \end{bmatrix}, \quad \boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \end{aligned} \tag{14}$$

An approximate solution to equation (13) can be found as

$$\begin{aligned} \boldsymbol{\tau} &= \mathbf{G}(\bar{\boldsymbol{\tau}})\bar{I}(\bar{\boldsymbol{\tau}}) \\ \mathbf{G}(\bar{\boldsymbol{\tau}}) &= \left(\mathbf{H}^T(\bar{\boldsymbol{\tau}})\mathbf{H}(\bar{\boldsymbol{\tau}}) + \mathbb{I}_2\right)^{-1}\mathbf{H}^T(\bar{\boldsymbol{\tau}}) \end{aligned} \tag{15}$$

where $\mathbb{I}_2$ is the 2×2 identity matrix.

4.2 The Iterative Solution

Define $\overline{\boldsymbol{\tau}}(k)$ to be the accumulated shifts from initial step until the kth step

$$\overline{\boldsymbol{\tau}}(k) = \sum_{i=0}^{k} \boldsymbol{\tau}(i) \tag{16}$$

Starting with an initial value $\overline{\boldsymbol{\tau}}(0) = 0$, the linear equation (13) and its solution (15) can be used in an iterative manner as follows

$$\begin{aligned}
\overline{\boldsymbol{\tau}}(0) &= 0 \\
e(0) &= \overline{I}(\overline{\boldsymbol{\tau}}(0)) \\
\boldsymbol{\tau}(1) &= \mathbf{G}(\overline{\boldsymbol{\tau}}(0))e(0) \\
&= \mathbf{G}(\overline{\boldsymbol{\tau}}(0))\overline{I}(\overline{\boldsymbol{\tau}}(0)) \\
e(1) &= \overline{I}(\overline{\boldsymbol{\tau}}(1)) - \mathbf{H}(\overline{\boldsymbol{\tau}}(1))[\boldsymbol{\tau}(0) + \boldsymbol{\tau}(1)] \\
&= \overline{I}(\overline{\boldsymbol{\tau}}(1)) - \mathbf{H}(\overline{\boldsymbol{\tau}}(1))\overline{\boldsymbol{\tau}}(1) \\
\boldsymbol{\tau}(2) &= \mathbf{G}(\overline{\boldsymbol{\tau}}(1))e(1) \\
&= \mathbf{G}(\overline{\boldsymbol{\tau}}(1))\left[\overline{I}(\overline{\boldsymbol{\tau}}(1)) - \mathbf{H}(\overline{\boldsymbol{\tau}}(1))\overline{\boldsymbol{\tau}}(1)\right] \\
&\vdots
\end{aligned} \tag{17}$$

In general

$$\boldsymbol{\tau}(k+1) = \mathbf{G}(\overline{\boldsymbol{\tau}}(k))\left[\overline{I}(\overline{\boldsymbol{\tau}}(k)) - \mathbf{H}(\overline{\boldsymbol{\tau}}(k))\overline{\boldsymbol{\tau}}(k)\right] \tag{18}$$

Substituting $\overline{I}(\overline{\boldsymbol{\tau}}(k))$ from equation (14) into equation (18) yields the formula for the iterative numerical solution as

$$\boldsymbol{\tau}(k+1) = \mathbf{G}(\overline{\boldsymbol{\tau}}(k))\left[I_2 - \mathbf{u}^T(\overline{\tau}_1(k))\mathbf{A}\mathbf{u}(\overline{\tau}_2(k))\right] \tag{19}$$

When $\boldsymbol{\tau}(k+1)$ in the iterative equation (21) converges to zero (or a *small enough* number near zero) we get $\overline{\boldsymbol{\tau}}(k+1)$ such that

$$I_2 \simeq \mathbf{u}^T(\overline{\tau}_1(k+1))\mathbf{A}\mathbf{u}(\overline{\tau}_2(k+1)) \tag{20}$$

which is the solution to both problems of motion estimation and inverse polynomial interpolation. Finally, algebraic manipulation of (19) and using (16) simplify the solution into the iterative formula given by

$$\overline{\boldsymbol{\tau}}(k+1) = \overline{\boldsymbol{\tau}}(k) + \frac{I_2 - \mathbf{u}^T(\overline{\tau}_1(k))\mathbf{A}\mathbf{u}(\overline{\tau}_2(k))}{\mathbf{H}(\overline{\boldsymbol{\tau}}(k))\mathbf{H}(\overline{\boldsymbol{\tau}}(k))^T}\mathbf{H}(\overline{\boldsymbol{\tau}}(k))^T \tag{21}$$

Our solution in (21) is closely related to the algorithm proposed by (Biemond et al., 1987). The major difference is that in (Biemond et al., 1987) a bilinear interpolation was used to calculate the displacement frame difference, and the spatial gradients were obtained by rounding off the displacement estimates; whereas in we use polynomial interpolation which provides better interpolation and simplifies calculating the gradients. Also, (Biemond et al., 1987) used observations from a block of pixels.

Motion estimation results can be improved significantly by testing multiple initial values. Figure 1 shows the chosen initial positions for the 5×5 neighborhood (i.e. $p = 2$). For the $(2p+1) \times (2p+1)$ neighborhood, the number of initial values $\overline{\boldsymbol{\tau}}(0)$ is p^2. The different initial values are sorted and tried according to their distance from the mean shift obtained for the previously processed adjacent pixels in I_2, starting with the closest. This also establishes dependency between the motions of the image pixels. For an initial value, if the iterative equation (21) converges to a solution before reaching a specified maximum number of iterations, the result is recorded and there would be no need to try the other initial values. Otherwise, the next initial value is tried.

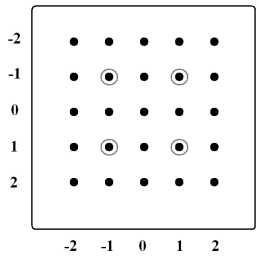


Figure 1: The circled pixel positions are the chosen initial positions for the 5×5 neighborhood.

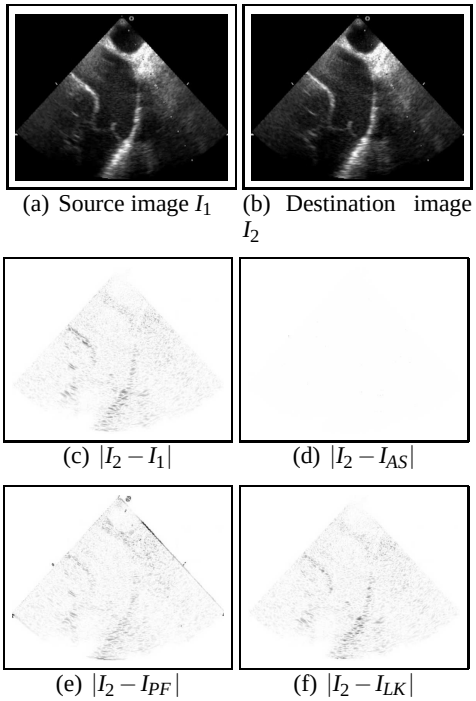
5 RESULTS

We tested our method using gray-scaled images. For comparison, motion in the same frames was estimated by the elastic image registration method by (Periaswamy and Farid, 2003) and the widely-used optical flow method by (Lucas and Kanade, 1981). The Matlab code for Periaswamy and Farid's method is available on the internet (Web, 1). Examples show that our method provides better performance.

In the first example (Figure 2) two images are extracted from an echocardiography video (Web, 2). The images are of size 430×550 pixels. The second example (Figure 3) shows two images extracted from a video recorded during a robotic-assisted repair of a pulmonary artery (Web, 3). The images are of size 240×352 pixels. For both examples we chose $p = 7$, a convergence threshold of 0.001 and the maximum number of iterations to be 20.

For each example, we computed the peak signal-to-noise ratio (PSNR) for the displaced frame difference. The PSNR equation is defined by (22) and the results are listed in Table 1.

$$PSRN = 10\log_{10}\frac{255^2 RC}{\sum_{r=1}^{R}\sum_{c=1}^{C}\left(I_2(r,c) - I_s(r,c)\right)^2} \tag{22}$$

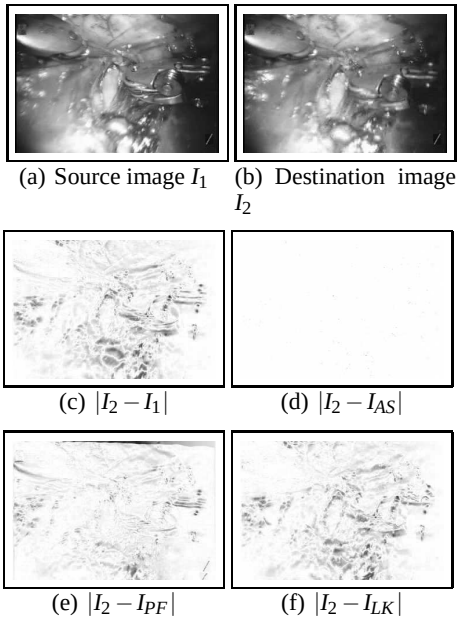


(a) Source image I_1 (b) Destination image I_2

(c) $|I_2 - I_1|$ (d) $|I_2 - I_{AS}|$

(e) $|I_2 - I_{PF}|$ (f) $|I_2 - I_{LK}|$

Figure 2: Motion estimation between two successive frames from echocardiography video.



(a) Source image I_1 (b) Destination image I_2

(c) $|I_2 - I_1|$ (d) $|I_2 - I_{AS}|$

(e) $|I_2 - I_{PF}|$ (f) $|I_2 - I_{LK}|$

Figure 3: Motion estimation between two successive frames from robot-assisted artery surgery video.

Table 1: PSNR of displaced frame difference.

	Echocardiography	Artery Surgery
Our method	50.27 dB	42.34 dB
Periaswamy-Farid	25.66 dB	19.85 dB
Lucas-Kanade	27.10 dB	19.12 dB

In (Periaswamy and Farid, 2003) the motion within a small neighborhood was modeled locally by an affine transform. In video sequences the considered neighborhood may contain one or more different motions in addition to the stationary background. An attempt to model these motions and the static background using one affine transform will produce estimation errors. Our method does not suffer from this shortcoming because it works on a pixel level.

ACKNOWLEDGEMENTS

This work was supported by National ICT Australia and the Australian Research Council.

REFERENCES

Auvray, V., Bouthemy, P., and Lienard, J. (2006). Motion estimation in x-ray image sequences with bi-distributed transparency. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1057–1060.

Biemond, J., Looijenga, L., Boekee, D., and Plompen, R. (1987). A pel-recursive wiener-based displacement estimation algorithm. *Signal Processing*, pages 399–412.

Hemmendorff, M. (2001). *Motion estimation and compensation in medical imaging*. PhD Thesis, Linkopings University, Sweden.

Kurabayashi, Y., Kagei, S., Gotoh, T., and Iwasawa, T. (2005). Motion estimation for sequential medical images using a deformable model. *Systems and Computers in Japan*, 36:27–36.

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 121–130.

Olver, P. (2006). On multivariant interpolation. *Studies in Applied Mathematics*, 116:201–240.

Periaswamy, S. and Farid, H. (2003). Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 22:865–874.

Web (1). http://www.cs.dartmouth.edu/farid/research/.

Web (2). http://www.cts.usc.edu/videos.html.

Web (3). http://www.manbit.com/ers/ersvideo.asp.

ELASTIC IMAGE WARPING USING A NEW RADIAL BASIC FUNCTION WITH COMPACT SUPPORT

Zhixiong Zhang and Xuan Yang
College of Information Engineering, Shenzhen University, GuangDong province, 518060 China
terry_kim@21cn.com, xyang0520@263.com

Keywords: Image registration, Compact support, Elastic registration, Bending energy, Biharmonic equation.

Abstract: Thin plate spline (TPS) and compact support radial basis functions (CSRBF) are well-known and successful tools in medical image elastic registration base on landmark. TPS minimizes the bending energy of the whole image. However, in real application, such scheme would deform the image globally when deformation is local. Although CSRBF can limit the effect of the deformation locally, it cost more bending energy which means more information was lost. A new radial basic function named 'Compact Support Thin Plate Spline Radial Basic Function' (CSTPF) has been proposed in this paper. It costs less bending energy than CSRBF in deforming image locally and its global deformation effect is similar to TPS. Numerous experimental results show that CSTPF performs outstanding in both global and local image deformation.

1 INTRODUCTION

Elastic image registration is a significant content in medical image registration. And image deformation plays an important part in elastic image registration. The use of TPS for point-based elastic registration was first proposed by Bookstein (Bookstein, 1989). TPS forces the corresponding landmark to match each other exactly and minimizes the utilization of the bending energy of the whole image, therefore, it is widely used in various fields (Brown and Rusinkiewicz, 2004). However, the deformation of TPS is global, it would be problematic when only local difference exists (Ruprecht and Müller, 1993). N. Arad, D. Reisfeld(Arad and Reisfeld, 1995) has investigated Gaussian radial basis function(RBF) which reduces the global influence. And M. Fornefett, H.S.Stiehl (Fornefett et al.,1999),(Fornefett et al., 2001) used compact support radial basis function(CSRBF) in medical image registration.

Although TPS can minimize the bending energy of the whole image, it can not deform the image locally. CSRBF can deform the image locally; however, it costs much bending energy which means that the warping loses lots of information form original image. This weakness is especially distinct while the deformation is globally.

In this paper, we proposed a new compact support radial basic function to deform the elastic image. This function not only limits the image's deformation in a local domain, but also is a fundamental solution to the biharmonic equation . Its bending costs less energy consequently. Simultaneously, when the support set is wide, its warp effect is similar to TPS. Therefore, this new compact support radial basic function performs well in the local and global registration experiments.

2 THE LIMITATION OF ELASTIC IMAGE REGISTRATION PRESENTLY

TPS models the deformations by interpolating displacements between source and target points. The basic function of TPS is $U(r_i) = r_i^2 \log r_i^2$, r_i is the distance form the cartesian origin. TPS's basic function is a so-called fundamental solution to the biharmonic equation (Bookstein, 1989), which can minimize the bending energy(1).

$$E_{TPS}(f) = \iint \left| \left| \frac{\partial^2 f}{\partial x^2} \right|^2 + \left| \frac{\partial^2 f}{\partial x \partial y} \right|^2 + \left| \frac{\partial^2 f}{\partial y^2} \right|^2 \right| dxdy \qquad (1)$$

Local elastic image registration which bases on RBFs has the same interpolation function as TPS. It can not eliminate the global effect of the deformation. M. Fornefett and K. Rohr(Fornefett et

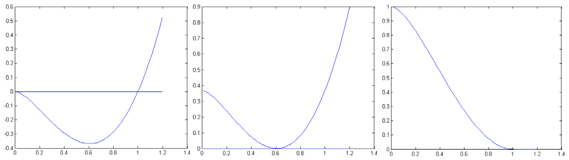
al.,1999), (Fornefett et al., 2001) applied Ψ - function of Wendland as RBFs for elastic registration of medical image. These radial basic functions have compact support.

These compact support radial basic functions can limit the deformation in a local domain. However, they are not the fundamental solution to the biharmonic equation. So they cost more bending energy in deforming the image, especially when their support radius are enormous, the loss of information of the source image is considerable.

3 COMPACT SUPPORT THIN PLATE SPLINE RADIAL BASIC FUNCTION

3.1 Compact Support Thin Plate Spline Radial Basic Function

In this paper, we aim to find out a function that not only is the solution to the biharmonic equation which can limit the bending energy, but have the characteristic of the functions which can deform image locally as well. Therefore, we use TPS's basic function: $U(r_i) = r_i^2 \log r_i^2$ to construct a new compact support radial basic function:



(a) (b) (c)

Figure 1: Compact support thin plate spline radial basic function : (a) $U(r_i) = r_i^2 \log r_i^2$ (b) fig.(a) displace 1/e upward (c) CSTPF.

As proven in figure.1 (a), TPS's original basic function $U(r_i) = r_i^2 \log r_i^2$ decreases at $0-1/\sqrt{e}$, after $1/\sqrt{e}$, the function increases rapidly afterwards. It can be noticed that the decreasing part ($0-1/\sqrt{e}$ part) of this function is similar to compact support radial basic functions which can deform the image locally. So we add a constant (constant value is the min value of this function: $1/e$) to this function (View at fig.1 (b)), and let its increasing part become zero. Then the function becomes:

$$U(r) = \begin{cases} r^2 \log r^2 + 1/e & , \quad r \le r_0 \\ 0 & , \quad r > r_0 \end{cases} \qquad (2)$$

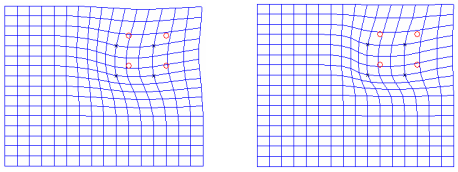
Figure.1 (c) shows that presently this function (2) has the characteristic of CSRBF. It has the max value at $r = 0$, it is a decreasing function which is compact supported meanwhile. Therefore, this function can be used for local deform interpolation. Furthermore, this function is the solution to the biharmonic equation and is able to decrease the bending energy. Because this function comes from TPS's original interpolate function, we name it as CSTPF (compact support thin plate spline radial basic function).

3.2 Local Deformation using CSTPF

For the purpose of further investigating elastic image deformation, we hypothesized that the source images have already been rigidly registered with the target image, ignoring the affine part of the interpolation function. In case of affine free deformation, the interpolation function is illustrated as follow: (This interpolation function was utilized in the following 2 chapters.)

$$f_x(x,y) = x + \sum_{i=1}^{n} w_{xi} U(|P_i - (x,y)|) \qquad (3)$$

First of all, we compared the performance of local deformation of CSTPF and CSRBF ($\psi_{a,3,2}$) at the same support set. Figure.2 explains the result of the deformation using the elastic registration approach base on CSTPF and CSRBF with four pairs of manual landmarks. It can be noticed that CSTPF is the solution to the biharmonic equation, correspondingly, less bending energy is required than using CSRBF, which means source image's information is better saved and the deformation has been greatly improved.



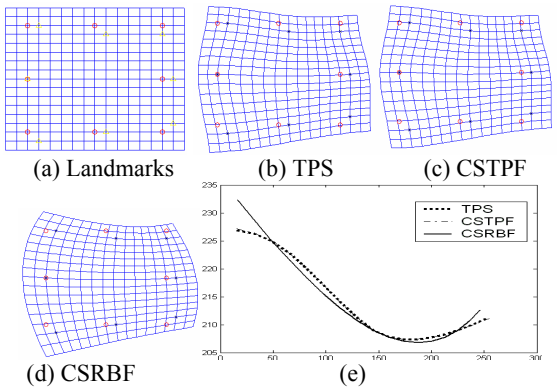
(a) CSTPF (b) CSRBF

Figure 2: Local deformation results (support radius $r = 100$) (a)Deformation using CSTPF, the cost of bending energy is3.342; (b) Deformation using CSRBF, the cost of bending energy is 6.323.

3.3 Global Deformation using CSTPF

It has been proved that CSFPF preformed well in local deformation in last paragraph, now let's discuss how it perform in global deformation.

Figure.3 is a contrast of the global deformations using the elastic registration approach base on TPS 、CSTPF、CSRBF with manual landmarks. In this Figure shows we can see that the deformations using the elastic registration approach base on CSTPF (Fig.3 (b)) and TPS (Fig.3 (c)) are almost the same. Figure.3(e) takes one line out of the deformation's results and makes a comparison. It is shown that the deformation's lines of CSTPF and TPS are almost superposed. This result illustrated that image deformation using CSTPF can keep the advantage of TPS in global deformation, which can not be achieved by using CSRBF.



(a) Landmarks (b) TPS (c) CSTPF

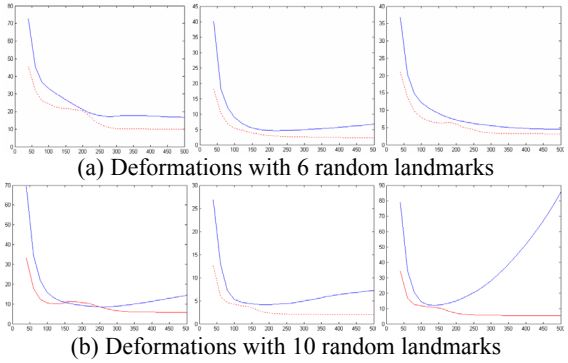
(d) CSRBF (e)

Figure 3: Global deformation contrast (support radius $r = 10000$) : (a) Landmarks' position; (b) global deformations using TPS; (c) global deformations using CSTPF; (d) global deformations using CSRBF; Contrast of the third line of Figure 3 (b)(c)(d), Notice Figure 3 (b) and (c) are almost superposed.

Experimental results have proved that image deformation using the elastic registration approach base on CSTPF is better than CSRBF.

To better illuminate the problem and aiming to compare the bending energy cost at different support radii, we experimented on 6 groups of deformations with random landmarks using the elastic registration approach based on CSTPF and CSRBF. Figure.4 shows the deformations' bending energy cost at different radii. In this graph, it is evident that image deformation using CSTPF costs less bending energy than CSRBF.

Consequently, the analysis and experiments in this chapter indicate that image global deformation using the elastic registration approach bases on CSTPF is similar to those on TPS. Furthermore, it is capable to localize the image deformation domain while TPS can not. In local image deformation, utilization of the elastic registration approach bases on CSTPF costs less bending energy than CSRBF with the same support radius.



(a) Deformations with 6 random landmarks

(b) Deformations with 10 random landmarks

Figure 4: Contrast of deformations' bending energy cost with random landmarks usizng CSTPF and CSRBF (X-axis: support radius, Y-axis is deformation's bending energy cost. real line in figure: CSRBF's energy cost , broken line in figure: CSTPF's energy cost): (a) Deformations with 6 random landmarks; (b) Deformations with 10 random landmarks.

4 EXPERIMENTAL RESULTS OF MEDICAL IMAGES

In this chapter, we have prepared two experiments in practical situations. With manual landmark, we compared the registration results for medical images using the elastic registration approach base on CSTPF and CSRBF.

In Figure 5, we can compare the results of deformation using CSTPF and CSRBF. They look similar but definitely not the same. Observing their edge comparison (Figure 5 (d) and (e)), it is revealed that after deformation, figure 5(d) has more edge information than figure 5(e) (as shown by the arrowhead), which means more information was saved by using CSTPF than CSRBF.

Finally, we employed another experiment to demonstrate that global deformation using CSTPF is better than CSRBF. In this experiment, we used an image of deferent mode, figure.6 (a) is MRI image and figure 6 (b) is CT image. It can be easily noticed that the source image and target image are just the same as they have no deformation. However, because we get landmarks manually, it is liable to have some artificial errors which are, however, considered as allowable errors. Given that these

allowable errors are unavoidable, the source image deform globally.

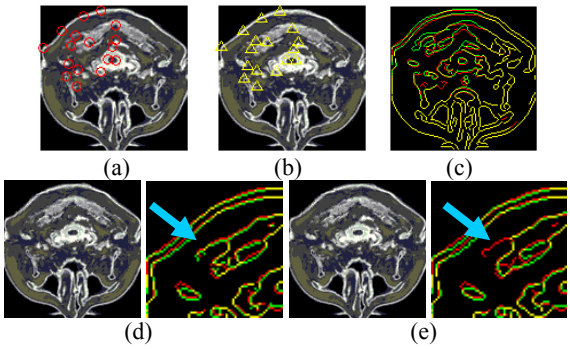


Figure 5: Comparison of local deformation using CSTPF and CSRBF($\psi_{a,3,2}$) ($r = 40$): (a) Source Image(with landmarks); (b) Target Image (with landmarks); (c) Comparison of edge of (a) and (b); (d)Deformation using CSTPF and the edge comparison of deformed image and original image(top left corner); (e)Deformation using CSTPF and the edge comparison of deformed image and original image(top left corner).

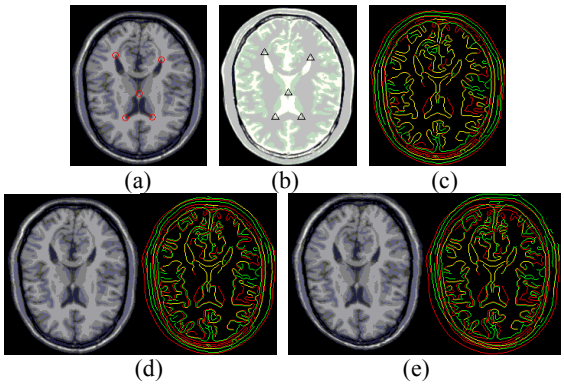


Figure 6: Comparison of global deformation using CSTPF and CSRBF($\psi_{a,3,2}$) ($r = 1000$):(a) Source Image(with landmarks); (b) Target Image (with landmarks); (c) Comparison of edge of (a) and (b); (d)Deformation using CSTPF and the edge comparison of deformed image and original image(global), the deformation cost 0.051327 bending-energy; (e)Deformation using CSTPF and the edge comparison of deformed image and original image(global), the deformation cost 0.19555 bending-energy.

It can be seen in figure.6 (d) that the image after global deformation using CSRBF has changed its shape significantly, while CSTPF kept the shape of source image satisfied. This demonstrated that CSTPF has stronger capability in global deformation than CSRBF.

5 CONCLUSIONS

In conclusion, TPS performs better in image global deformation, but it is not suitable for the local elastic registration. CSRBF can be used in the local registration, but it cost more bending energy which means that it will lose more information during the deformation. Moreover, its global deformation is not as well as TPS. In this paper, we proposed a new radial basic function 'CSTPF' which is a solution to the biharmonic equation. In local deformation, using CSTPF cost less bending energy. In global deformation, using CSTPF can keep the topology of source image. Additionally, it can save the information more integrated and it is approaches to TPS. Hence CSTPF is considered superior in image deformation.

ACKNOWLEDGEMENTS

This paper is sponsored by Nation Natural Science Funds (NO 60572101) of China.

REFERENCES

Fred L. Bookstein, Principal warps: thin-plate splines and the decomposition of deformations, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 11, No.6, pp.567-585, June 1989.

B. J. Brown, S. Rusinkiewicz, Non-rigid range-scan alignment using thin-plate splines, 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on Sept. 2004 , pp. 759- 765.

D. Ruprecht, H. Müller, Free form deformation with scattered data interpolation mehtods, Computing Supplementum 8 (1993), pp. 267-281.

N. Arad, D. Reisfeld, Image warping using few anchor points and radial functions, Comp. Graphics Forum, 14(1): pp. 35-46, 1995.

M. Fornefett, K. Rohr, H. S. Stiehl, Elastic registration of medical images using radial basis functions with compact support, CVPR'99, IEEE Computer Society PR00149, Fort Collins, CO, USA, 23-25 June 1999, pp.402-407

M. Fornefett, K. Rohr, H. S. Stiehl, Radial basis functions with compact support for elastic registration of medical images, Image Vision Comput, 2001, vol. 19(1-2), pp.87-96

COMPARATIVE STUDY OF SEVERAL NOVEL ACOUSTIC FEATURES FOR SPEAKER RECOGNITION

*†Vladimir Pervouchine, *Graham Leedham, *‡Haishan Zhong, *David Cho and *†Haizhou Li
*Nanyang Technological University, School of Computer Engineering, N4 Nanyang Ave, Singapore 639798
assycho@ntu.edu.sg, g.leedham@gmail.com
†Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
{vpervouchine,hli}@i2r.a-star.edu.sg
‡Panasonic Singapore Laboratories, 1022 Tai Seng Avenue, #06-3530 Tai Seng Ind Est, Singapore 534415
haishan.zhong@sg.panasonic.com

Keywords: Speaker recognition, feature extraction, feature evaluation.

Abstract: Finding good features that represent speaker identity is an important problem in speaker recognition area. Recently a number of novel acoustic features have been proposed for speaker recognition. The researchers use different data sets and sometimes different classifiers to evaluate the features and compare them to the baselines such as MFCC or LPCC. However, due to different experimental conditions direct comparison of those features to each other is difficult or impossible. This paper presents a study of five new recently proposed acoustic features using the same data (NIST 2001 SRE), and the same UBM-GMM classifier. The results are presented as DET curves with equal error ratios indicated. Also, an SVM-based combination of GMM scores produced on different features has been made to determine if the new features carry any complimentary information. The results for different features as well as for their combinations are directly comparable to each other and to those obtained with the baseline MFCC features.

1 INTRODUCTION

Speaker recognition is the process of automatic identification or verification of a speaker using the information obtained from his/her speech. Verification permits access control by voice as well as facilitates in crime investigation if recordings of phone conversations are available. Speaker recognition systems includes speaker identification and speaker verification; however, studies usually focus on speaker verification only. Text-independent verification has gained much attention because does not require a user to speak any pre-defined phrases for the system to operate and thus is an attractive method of personal verification.

To represent a speaker features extracted from the audio file are used. It is supposed that different speakers can be represented differently in the feature space. Therefore by building statistical models that approximate the distribution of feature vectors for different speakers, the conditional probability of the speaker being who he claims can be estimated.

Finding good features with low intra-speaker variation and high inter-speaker variation, as well as not too sensitive to channel type, is an important problem in speaker recognition. Commonly used features are Linear Prediction (LP) based features such as Linear Predictive Cepstral Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC) (Gish and Schmidt, 1994). Although recently research focus has been shifted mostly to developing methods of elimination of channel effects, a number of new and novel features have been proposed (Wang and Wang, 2005; Sant'Ana et al., 2006; Cordeiro and Ribeiro, 2006; Sri Rama Murty and Yegnanarayana, 2006). New classification method to be used with the new features has also been proposed in (Sant'Ana et al., 2006). Researchers use different data sets to compare performance of speaker verification systems with their features to that of the baseline systems. Therefore, it is often difficult or impossible to compare the effectiveness of the new features to each other and to baseline features (such as MFCC).

This paper presents a comparative study of five new features proposed in 2005–2006 publications. The features studied are *Mean Energy within Critical Bands (MECB)* and *Difference of Mean Energy within Critical Bands (DMECB)* (Wang and Wang, 2005), *pH* features based on Hurst parameter and fractional

Brownian motion model (Sant'Ana et al., 2006), *Mel Line Spectrum Frequencies (MLSF)* (Cordeiro and Ribeiro, 2006), and *Residual Phase* (Sri Rama Murty and Yegnanarayana, 2006). The paper is organised as follows. Section 2 shortly describes the features and section 3 presents the feature combination scheme. Section 4 discusses the experimental setup and the results, and section 5 draws the conclusions.

2 FEATURES STUDIED

Mel Frequency Cepstral Coefficients. Commonly used MFCC features were taken as the baseline. The data set audio files were divided into frames of 30 ms length with 1/3 overlap using Hamming window. Twelve MFCC coefficients were calculated for each frame along with their first and second differences (ΔMFCC and $\Delta\Delta$MFCC) resulting in 36-dimensional feature vectors. The feature values were normalised by subtracting the mean and dividing by the standard deviation.

Mel Line Spectrum Frequencies. Mel Line Spectrum Frequencies (MLSF) are similar to Line Spectrum Frequencies calculated from Linear Prediction (LP) coefficients. The difference is in taking an advantage of mel frequency warping, emphasising the information in lower frequencies (Cordeiro and Ribeiro, 2006). To extract MLSF features, the signal was divided into 30 ms frames windowed using Hamming window with 1/3 frame overlap. Fast Fourier Transform (FFT) and mel filter bank were used to generate mel spectrum. Then the inverse Fourier transform was applied to get the mel autocorrelation of the signal. The MLSF features were then calculated via the Levinson-Durbin recursion. An LP filter of order 16 was used resulting in 16-dimensional feature vectors. The feature values were normalised by subtracting the mean and dividing by the standard deviation. Since addition of the first and second differences (ΔMLSF and $\Delta\Delta$MLSF) might increase the verification accuracy both differences were calculated. In the original paper MLSF features were evaluated on NIST 2002 SRE database.

Residual Phase. A person's vocal tract can be modelled as an excitation source and a set of filters that characterise the vocal tract shape. While LP coefficients approximate the vocal tract shape, the excitation source can be evaluated from the residual signal:

$$e_n = s_n + \sum_{k=1}^{p} a_k s_{n-k} \tag{1}$$

where s_k is the signal, a_k are the LP coefficients. Examples of features calculated from the residual signal include Haar Octave Coefficients of Residue (HOCOR) (Zheng and Ching, 2004) and Residual Phase (Sri Rama Murty and Yegnanarayana, 2006). The latter was evaluated on NIST 2003 SRE database. To extract the residual phase, the analytic signal $R_n = r_n + jh_n$ is calculated from the residual signal r_n, where h_n is the Hilbert transform h_n of r_n. The phase is then calculated from the analytic signal as:

$$\theta_n = \arccos\left(r_n / \sqrt{r_n^2 + R_n^2}\right) \tag{2}$$

Authors (Sri Rama Murty and Yegnanarayana, 2006) recommend to calculate the residual phase from short segments of speech of around 5 ms, which is justified by the period of the bursts in the excitation source. In the our study the signal was divided into 6 ms frames with 1/3 overlap. LP of orders 6 and 10 were tried.

Hurst Parameter Features. Hurst parameter features were proposed for speaker recognition (Sant'Ana et al., 2006) and evaluated on BaseIME database developed in the Instituto Militar de Engenharia. The feature vector is a vector of Hurst parameters calculated for frames of a speech signal via Abry-Veitch Estimator using discrete wavelet transform. To extract the features the speech signal was divided into 80 ms frames with 1/2 overlap, which were chosen to make the extraction similar to that presented in the original paper. Daubechies wavelets with 4, 6, and 12 coefficients were used, thus resulting in pH_4, pH_6, and pH_{12} features. The depth of wavelet decomposition was to be 5, 4, and 3 for pH_4, pH_6, and pH_{12} respectively.

Mean Energy within Critical Bands. Fractional Fourier transform (FrFT) is a generalisation of the ordinary (integer) Fourier transform. Mean Energy within Critical Bands (MECB) features based on the fractional Fourier transform were proposed in (Wang and Wang, 2005) and evaluated on a custom data set. The critical bands are formed by warping frequency according to the mel or bark scale. MECB_p are calculated by taking the fractional Fourier transform of order p of each frame of the signal. For i-th critical band $f_i \ldots f_{i+1}$ the log of mean energy is

$$E_i = \log \frac{\int_{f_i}^{f_{i+1}} |F(f)|^2 \, df}{f_{i+1} - f_i} \tag{3}$$

For two MECB features of orders p_1 and p_2 the difference MECB (DMECB) features are calculated as:

$$\mathrm{DMECB}_{p_1 - p_2} = \mathrm{MECB}_{p_1} - \mathrm{MECB}_{p_2} \tag{4}$$

In our study the signal was divided into 30 ms long frames with 1/3 overlap. MECB of orders $p = 0.5, 0.6, \ldots, 1.0$ were extracted. DMECB were calculated for a fixed p_1 of 1.0 and p_2 $0.5 \ldots 0.9$.

3 COMBINATION OF FEATURES

Combining different acoustic features can be performed in a number of ways. One way is to concatenate the feature vectors of the corresponding frames. However, this leads to feature vectors of very high dimensionality, which means much more data is required for reliable training of a classifier. Thus the concatenation was only done for low-dimensional feature vectors pH, while for the high-dimensional features another method was used. A GMM enables modelling the conditional probability density functions in the feature space for each class. A GMM classifier returns a score for each given pattern, which is an estimation of the log likelihood ratio for the hypothesis that the speaker is who he claims to be (Reynolds and Rose, 1995). These scores from GMM classifiers for each of the acoustic features were used as features. The resulting score feature vectors were used with an SVM classifier.

4 EXPERIMENTS AND RESULTS

All experiments were conducted on NIST 2001 Speaker Recognition Evaluation (SRE) database, single-speaker files. The audio files sampled at 8 kHz were pre-emphasised with filter coefficient of 0.97 and divided into frames as described above. For all features a Gaussian Mixture Model (GMM) classifier of 512 multivariate normal distributions with diagonal covariance matrices was used (Reynolds and Rose, 1995). The Universal Background Models (UBM) were trained on samples from 82 male and 56 female speakers. The resulting Detection Error Tradeoff (DET) curves and the Equal Error Ratios (EER) are shown in Fig. 1(a)–(g).

Individual Features. The results achieved with MFCC features with the first and second differences were taken as the baseline (Fig. 1(a)). As seen from the DET curves in Fig. 1(b), adding the first difference to MLSF improves the speaker verification accuracy, which is in agreement with the results in (Cordeiro and Ribeiro, 2006). Adding the second difference improves the accuracy further. Because of high dimensionality of the resulting feature vectors (48) more training data may lead to better system performance.

Fig. 1(c) shows the DET curves for Residual Phase features and two different order LP filters. The difference in the LP filter order does not result in a significant difference in the speaker verification accuracy. It was also found that adding the first difference features does not change the system performance either, so the second difference was not tried.

Features pH_{4+6+12} were obtained by concatenating feature vectors pH_4, pH_6, and pH_{12} for each frame. It was found that performance of the speaker verification system is similar when either one of pH_4, pH_6, pH_{12} are used. Concatenating them into 12-dimensional pH_{4+6+12} vectors leads to a dramatic improvement in the accuracy with EER dropping from 29.0% to 20.8% (Fig. 1(d)).

The accuracy of speaker verification for MECB_p features declines with p of FrFT (Fig. 1(e)). This is in accordance with the results reported in (Wang and Wang, 2005), while the results for $\mathrm{DMECB}_{1.0-p_2}$ features with various p_2 (Fig. 1(f)) are different from that reported in the paper: the highest speaker verification accuracy was achieved for $p_2 = 0.5$ and for $p_2 = 0.6 \ldots 0.9$ the accuracy decreased with increase of p_2. Adding the difference features to MECB and DMECB did not lead to accuracy improvement.

Table 1: Equal error rates for MECB features of different orders.

MECB_p, p	1.0	0.9	0.8	0.7	0.6	0.5
EER, %	17.6	18.7	21.2	24.2	27.5	31.4

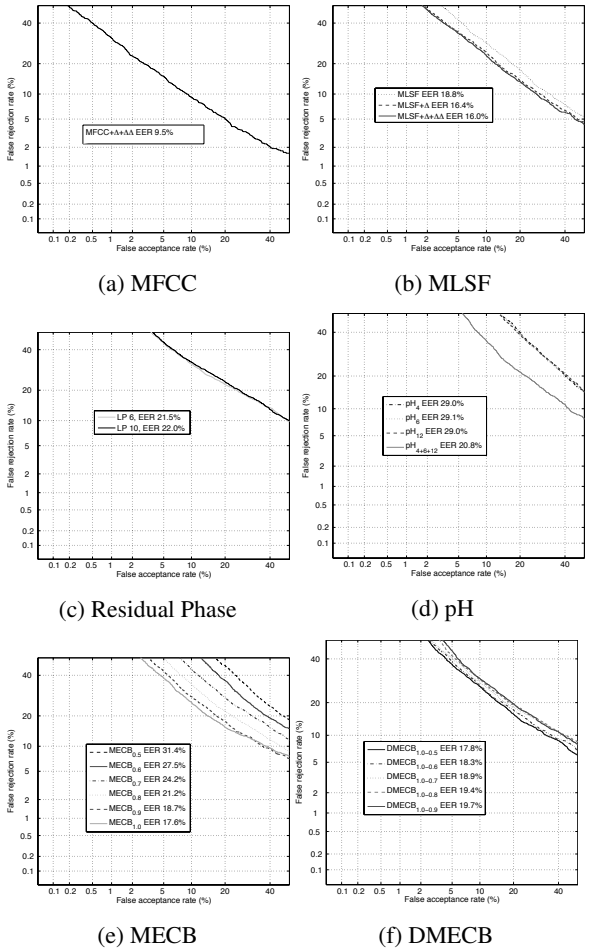
Table 2: Equal error rates for DMECB features of different orders.

$\mathrm{DMECB}_{1.0-p_2}$, p_2	0.9	0.8	0.7	0.6	0.5
EER, %	19.7	19.4	18.9	18.3	17.8

Table 3: Summary of equal error rates for different feature types and their SVM combination.

Feature type	**EER, %**	**Feature type**	**EER, %**
MFCC+Δ+ΔΔ	9.5	Residual phase	21.5
MLSF+Δ+ΔΔ	16.0	pH_{4+6+12}	20.8
$\mathrm{MECB}_{1.0}$	17.6	$\mathrm{DMECB}_{1.0-0.5}$	17.8
Combined	8.7		

Combination of Features. To make the results comparable to those of acoustic features alone a 5-fold cross-validation scheme was applied. The test set of speakers was divided into 5 approximately equal parts. Every time one different part was left for testing and four others were used for training the SVM, resulting in 5 experiments in total. The SVM was designed to produce a soft decision, which



(a) MFCC (b) MLSF

(c) Residual Phase (d) pH

(e) MECB (f) DMECB

Figure 1: DET curves for different feature types.

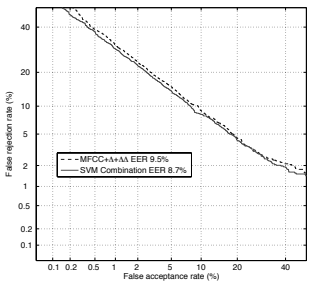


Figure 2: DET curve for combination of acoustic features.

was then treated as a score for plotting the DET curve and calculating the EER. It was decided to combine MFCC+Δ + $\Delta\Delta$, MLSF+Δ + $\Delta\Delta$, Residual Phase with LP of order 6, pH_{4+6+12}, $\text{MECB}_{1.0}$, and $\text{DMECB}_{1.0-0.5}$, choosing one feature from each group with the best performance. The results of the combination are shown in Fig. 2 with the EER in Table 3 with the DET curve for the MFCC+Δ+$\Delta\Delta$ plotted for comparison.

5 CONCLUSIONS

As assessed on NIST 2001 SRE database, none of the novel acoustic features considered in this study outperformed the MFCC features. MLSF and MECB features have performance that is comparable to that of MFCC. Features pH showed a high accuracy of speaker verification taking into account their low dimensionality (5, 4, and 3 for pH_4, pH_6, and pH_{12} respectively), hence they may be attractive when limited training data is available. Combination of several different acoustic features resulted in significantly higher accuracy of speaker verification. We can conclude that the studied features carry additional information about speakers. How big is the contribution of each of the features into the speaker verification accuracy, however, needs to be established. To determine this either all possible feature combinations have to be tried, or feature selection methods have to be used.

REFERENCES

Cordeiro, H. and Ribeiro, C. (2006). Speaker characterization with MLSF. In *Odyssey 2006: The Speaker and Language Recognition Workshop*, San Juan, Puerto Rico.

Gish, H. and Schmidt, M. (1994). Text-independent speaker identification. *IEEE Signal Processing Magazine*, pages 18–32.

Reynolds, D. and Rose, R. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Speech Audio Process.*, 3(1):72–83.

Sant'Ana, R., Coehlo, R., and Alcaim, A. (2006). Text-independent speaker recognition based on the Hurst parameter and the multidimensional fractional Brownian motion model. *IEEE Trans. Acoust., Speech, Signal Process.*, 14(3):931–940.

Sri Rama Murty, K. and Yegnanarayana, B. (2006). Combining evidence from residual phase and MFCC features for speaker recognition. *IEEE Signal Process. Lett.*, 13(1):52–55.

Wang, J. and Wang, J. (2005). Speaker recognition using features derived from fractional Fourier transform. In *4th IEEE Workshop on Automatic Identification (AutoID 2005)*, Buffalo, NY, USA.

Zheng, N. H. and Ching, P. C. (2004). Using Haar transformed vocal source information for automatic speaker recognition. In *Proc. IEEE ICASSP*, volume 1, pages 77–80, Montreal, Canada.

COMBINING NOVEL ACOUSTIC FEATURES USING SVM TO DETECT SPEAKER CHANGING POINTS

*Haishan Zhong, *David Cho, *†Vladimir Pervouchine and *Graham Leedham
*Nanyang Technological University, School of Computer Engineering, N4 Nanyang Ave, Singapore 639798
zhonghaishan@gmail.com, assycho@ntu.edu.sg,g.leedham@gmail.com
†Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
vpervouchine@i2r.a-star.edu.sg

Keywords: Speaker recognition, Feature extraction, Feature evaluation.

Abstract: Automatic speaker change point detection separates different speakers from continuous speech signal by utilising the speaker characteristics. It is often a necessary step before using a speaker recognition system. Acoustic features of the speech signal such as Mel Frequency Cepstral Coefficients (MFCC) and Linear Prediction Cepstral Coefficients (LPCC) are commonly used to represent a speaker. However, the features are affected by speech content, environment, type of recording device, etc. So far, no features have been discovered, which values depend only on the speaker. In this paper four novel feature types proposed in recent journals and conference papers for speaker verification problem, are applied to the problem of speaker change point detection. The features are also used to form a combination scheme using an SVM classifier. The results shows that the proposed scheme improves the performance of speaker changing point detection as compared to the system that uses MFCC features only. Some of the novel features of low dimensionality give comparable speaker change point detection accuracy to the high-dimensional MFCC features.

1 INTRODUCTION

The aim of speaker changing point detection (speaker segmentation) is to find acoustic events within an audio stream (e.g finding the speaker changing point in the continuous speech files according to different speakers' characteristics). Automatic segmentation of an audio stream according to speaker identities and environmental conditions have gained increasing attention. Since some speech files are obtained from telephone conversations or recorded during meetings, there are more than one person speaking in the audio recordings. In such cases before performing speaker recognition it is necessary to separate audio signal according to different speakers. Features extracted from a speech waveform are used to represent the characteristics of the speech and speaker. Among the features acoustic features are those based on spectrograms of short-term speech segments. However, the feature values that represent a speaker also vary due to speech content, environment, type of recording device, etc. So far no features have been discovered, whose values only depend on the speaker. Also, different speech features contains different information about a speaker: some features reflect a person's vocal tract shape while others may characterise the vocal tract excitation source.

Generally there are three main techniques for detecting the speaker changing points: decoder guided, metric based and model based. In this paper, a method using Support Vector Machines (SVM) to find speaker changing point in a continuous audio file is presented. SVM is a binary classifier that constructs a decision boundary to separate the two classes. SVM has gained much attention since the experimental results indicate that it can achieve a generalisation performance that is greater than or equal to other classifiers, but requires less training data to achieve such an outcome (Wan and Campbell, 2000). Speaker segmentation can be treated as a binary decision task: the system must decide whether or not a speech frame has the speaker changing point. This study uses the SVM for seeking speaker changing points by combining commonly used acoustic features with several novel acoustic features proposed recently. The novel features have been recently proposed by different re-

searchers for the problem of speaker recognition.

The paper is organised as follows: section 2 describes the speaker segmentation method with Bayesian Information Criteria. In section 3 the feature extraction is described for each feature type. In section 4 the structure of SVM speaker segmentation is explained. Section 5 presents the experimental results and draws the conclusion.

2 SPEAKER SEGMENTATION WITH BIC

A speaker changing point detection algorithm using Bayesian Information Criterion (BIC) is proposed in (Chen and Gopalakrishnan, 1998). A speech signal is divided into partially overlapping frames of around 30 ms length using a Hamming window. Extraction of acoustic features is performed for each speech frame. A sliding window with minimum size W_{min} and maximum size W_{max} shifted by F frames is used to group several consecutive frames. For detail grouping algorithm the reader may refer to (Chen and Gopalakrishnan, 1998). Each segment contains a number of frames and is represented by the corresponding acoustic feature vectors. A segment can be modelled as a single Gaussian distribution. The distance between consecutive segments is calculated based on variances of the Gaussian distributions that model the segments in the feature space. The variance BIC (Nishida and Kawahara, 2003) was developed from BIC and used to represent the distance between two speech segments represented as their feature vectors. Variance BIC is formulated with the following function:

$$\begin{aligned}\Delta BIC^{i}_{variance} = &-\frac{n_1+n_2}{2} log_i|\Sigma_0| + \\ &+\frac{n_1}{2} log_i|\Sigma_1| + \frac{n_2}{2} log_i|\Sigma_2| + \\ &+\alpha\frac{1}{2}(d+\frac{1}{2}d(d+1))log(n_1+n_2)\end{aligned} \quad (1)$$

where Σ_0, Σ_1 and Σ_2 are the covariance values of the whole segment, the first segment and the second segment respectively, n_i is the number of frames for the i-th segment, and d is the dimensionality of the acoustic feature vectors. The larger the variance BIC of two segments is, the larger is the probability that there is a speaker changing point between these two segments. A sliding window is used to calculate the variance BIC value for the whole speech files (Chen and Gopalakrishnan, 1998). Local maxima in variance BIC values of the whole speech are marked as the speaker changing points.

When different acoustic features are used, there will be different variance BIC values generated for a speech file. These values can be used as features forming feature vectors to be used for determining speaker changing points. Fig. 1 shows the process of generating a variance BIC vector after acoustic feature extraction. After the feature extraction for each frame, the feature vector of each frame is used to calculate the variance BIC values.

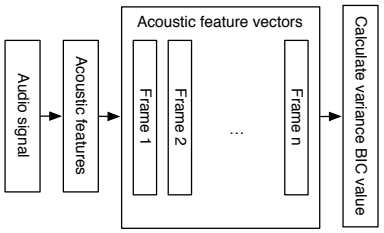


Figure 1: Generating variance BIC values for single type acoustic features.

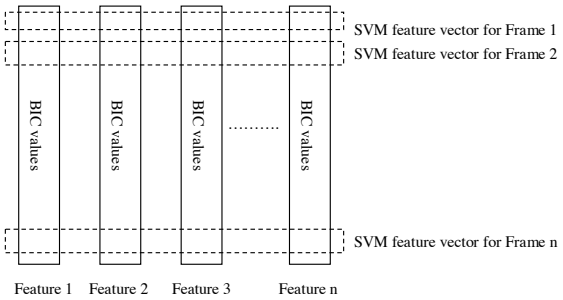


Figure 2: Combination of variance BIC values generated from different acoustic features into feature vectors.

Most of the acoustic features are of high dimensionality, and simple concatenation of the feature vectors will result in a feature vector of even higher dimensionality, which, in turn, will require too many training samples to be trained reliably. Instead, in the current study the variance BIC value is calculated from each of the acoustic features as described above. The BIC values calculated for the same frames using different acoustic features form new feature vectors are then used with the SVM classifier (Fig. 2).

3 EXTRACTION OF FEATURES

The features were extracted from speech sampled at 16 kHz. Mel Frequency Cepstral Coefficients (MFCC) (Oppenheim and Schafer, 2004) features were used to calculate the variance BIC values. MFCC vectors were extracted from 30 ms frames without overlap. The feature values were normalised by subtracting the mean and dividing by the standard

deviation. First order difference features were added. In addition, several novel acoustic features were used to calculate variance BIC values: Mel Line Spectrum Frequencies (MLSF) (Cordeiro and Ribeiro, 2006), Hurst parameter features (pH) (Sant'Ana et al., 2006), Haar Octave Coefficients of Residue (HOCOR) (Zheng and Ching, 2004), and features based on fractional Fourier transform (MFCCFrFT).

FrFTMFCC$_p$ are extracted similarly to MFCC with the only difference that the fractional Fourier transform of order p is used in place of the integer one. Features of various orders p were tried and FrFTMFCC$_{0.9}$ were chosen because they gave the next highest speaker segmentation accuracy after FrFTMFCC$_{1.0}$, which are the conventional MFCC, as measured by the F-score (see below).

MLSF are similar to Line Spectrum Frequencies calculated from LP coefficients and were proposed in the context of the speaker verification problem. A mel spectrum was generated via Fast Fourier Transform (FFT) and mel filter bank applied to 30 ms frames. The inverse Fourier transform was applied to calculate the mel autocorrelation of the signal, from which MLSF features were then calculated via Levinson-Durbin recursion. LP of order 10 was used. The feature values were normalised by subtracting the mean and dividing by the standard deviation.

Hurst parameter is calculated for frames of a speech signal via Abry-Veitch Estimator using discrete wavelet transform (Veith and Abry, 1998). In the current study a frame length of 60 ms was used, and Daubechies wavelets with four, six, and twelve coefficients were tried giving rise to pH$_4$, pH$_6$, and pH$_{12}$ features. The depth of wavelet decomposition was chosen to be 5, 4, and 3 for pH$_4$, pH$_6$, and pH$_{12}$ correspondingly, thus resulting in 5-, 4-, and 3-dimensional feature vectors (Sant'Ana et al., 2006).

While LP coefficients are aimed at characterising the person's vocal tract shape, information about the glottal excitation source can be extracted from the residual signal $e_n = s_n + \sum_{k=1}^{p} a_k s_{n-k}$. Haar Octave Coefficients of Residue (HOCOR) features are extracted by applying Haar transform to the residual signal. In the current study the LP of order 12 was applied to 30 ms frames. HOCOR$_\alpha$ features of order α 1, 2, 3, and 4 were extracted (Zheng and Ching, 2004).

4 SVM SPEAKER SEGMENTATION

Fig. 3 shows the structure of SVM speaker segmentation. To be used in SVM the frames which contain speaker changing point are labelled as -1, the frames without speaker changing point are labelled as 1. The acceptable error range of the found speaker changing points was chosen to be 1 second (Ajmera et al., 2004), which means the frames that are half a second before and after a speaker changing point are all labelled as -1. The variance BIC values that are obtained from different acoustic features are of different order. To use them as features in SVM a linear scaling is applied:

$$\hat{f}_i^j = \frac{f_i^j - \langle f_i \rangle}{\sigma_i} \tag{2}$$

where i represent different features, j is the frame number of the i-th feature, $\langle f_j \rangle$ is the mean value of f_j^j and σ_i is its standard deviation.

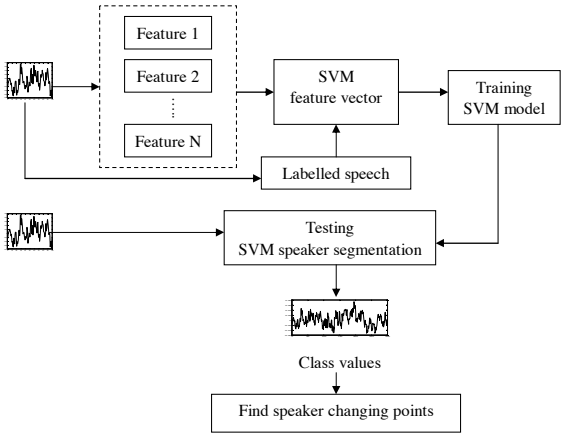


Figure 3: SVM speaker segmentation system.

The SVM classifier returns two values for each frame that are related to the distance to the separating hyperplane (either of them can be monotonically mapped into the conditional class probability). The values sum to one and indicate to what extent the frame belongs to class -1 (or 1). The -1 class value is analysed to determine the true speaker changing point. A peak search algorithm was used to determine the local maxima of -1 class value as we moved along the frames. The peak searching algorithm uses adaptive threshold in an attempt to eliminate small peaks due to noise and find only true local maxima.

5 RESULTS AND CONCLUSION

The NIST HUB-4E Broadcast News Evaluation data set was used in this study. The data was obtained from the audio component of a variety of television and broadcast news sources and each audio file consist of approximately one hour of speech in English

and includes the speech of several speakers in one audio channel (Hub, 1997). To evaluate the performance of the speaker changing point detection, two criteria were used: the precision of speaker changing points that were found and the number of missed changing points. The precision indicates the percentage of true turning points from the total number of turning points that were found. The recall indicates how many of the true turning points were missed. These two are combined into an F-score. F-score indicates how good a system is: it is high when both precision and recall values are high and low when either of them is low (Nishida and Kawahara, 2003).

Table 1: F-score, precision and recall for different features and their combination via SVM. d is the dimensionality of the acoustic feature vectors.

Feature	d	F-score	Precision	Recall
MFCC	26	0.62	0.61	0.62
MLSF	10	0.42	0.29	0.80
pH_4	5	0.52	0.67	0.43
pH_6	4	0.53	0.67	0.44
pH_{12}	3	0.55	0.68	0.46
$HOCOR_1$	6	0.42	0.54	0.35
$HOCOR_2$	5	0.37	0.47	0.30
$HOCOR_3$	4	0.31	0.39	0.26
$HOCOR_4$	3	0.30	0.38	0.25
$FrFTMFCC_{0.9}$	12	0.61	0.73	0.56
SVM_1	10	0.64	0.72	0.58
SVM_2	6	0.65	0.75	0.58

Table 1 (except for the two bottom rows) shows the speaker changing point detection results achieved when different acoustic features were used to calculate the variance BIC and the peak detection algorithm was used to detect speaker changing points from the BIC values. It is worth noticing that using pH features gives F-scores comparable to those when MFCC features are used, even though the dimensionality of feature vectors of pH features is far less than those of MFCC. This suggests that pH features may be a better choice when the training data set is small.

The features used for SVM combination 1 (SVM_1) are the 10 variance BIC values resulted from the 10 acoustic features. The results in Table 1 show that the proposed SVM speaker changing point detection scheme improves the speaker changing point detection performance as compared to each of the individual acoustic features, with a higher F-score of 0.64. This means that other acoustic features, which were originally proposed for speaker recognition problem, can be used for the problem of speaker segmentation as well. Because of low both precision and recall values achieved on HOCOR features, a combination of the acoustic features was attempted without HOCOR features. The results (SVM_2 in Table 1) were comparable with those of SVM_1. However, elimination of any other acoustic features from the combination degraded the speaker segmentation performance.

This study demonstrates that the new features do carry additional information about speaker differences to MFCC features, and some of them also have attractiveness because of their low dimensionality. Further study may find better ways of how to integrate complimentary information about speaker differences contained in the new features with traditional features such as MFCC and LPCC.

REFERENCES

(1997). NIST HUB-4E Broadcast News Evaluation.

Ajmera, J., McCowan, I., and Bourlard, H. (2004). Robust speaker change detection. *IEEE Signal Process. Lett.*, 11(8).

Chen, S. and Gopalakrishnan, P. (1998). Speaker, environment and channel change detection and clustering via the Bayesian Information Criterion. In *DARPA Speech Recognition Workshop*, pages 127–132.

Cordeiro, H. and Ribeiro, C. (2006). Speaker characterization with MLSF. In *Odyssey 2006: The Speaker and Language Recognition Workshop*, San Juan, Puerto Rico.

Nishida, M. and Kawahara, T. (2003). Unsupervised speaker indexing using speaker model selection based on Baysian Information Criterion. In *Proc. IEEE ICASSP*, volume 1, pages 172–175.

Oppenheim, A. and Schafer, R. (2004). From frequency to quefrency: a history of the cepstrum. *Signal Processing Magazine, IEEE*, (5):95–106.

Sant'Ana, R., Coehlo, R., and Alcaim, A. (2006). Text-independent speaker recognition based on the Hurst parameter and the multidimensional fractional Brownian motion model. *IEEE Trans. Acoust., Speech, Signal Process.*, 14(3):931–940.

Veith, D. and Abry, P. (1998). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inf. Theory*, 45(3):878–897.

Wan, V. and Campbell, M. (2000). Support vector machines for speaker verification and identification. pages 775–784.

Zheng, N. and Ching, P. (2004). Using Haar transformed vocal source information for automatic speaker recognition. In *IEEE ICASSP*, pages 77–80, Montreal, Canada.

A IMAGE PROCESSING METHOD FOR COMPARISON OF MULTIPLE RADIOGRAPHS

Chen Sheng, Li Li and Wang Pei
College of Mathematics and Science, Shanghai Normal University, Guilin Road 100, Shanghai, China
chnshn@hotmail.com

Keywords: Image Processing, Look up table (LUT), Computer-aided diagnosis (CAD).

Abstract: Portable chest radiography is the most commonly ordered radiographic test in the intensive care unit (ICU). In the ICU, a succession of portable images is usually taken over a period of time to monitor the progress of a patient's condition. A prompt diagnosis of any changes in the conditions of these ICU patients allows clinicians to provide immediate attention and treatments that are required to prevent the conditions from worsening and which could result in a treat to the patient's life. However, because of differences in X-ray exposure setting, patient and apparatus positioning, scattering, and grid application, for example, differences in image quality from on image to the next taken at different times can be significant. The differences in image quality make it difficult for clinicians to compare images to detect subtle changes. This paper presents an image-rendering method that reduces the variability in image appearance and enhances the diagnostic quality of these images. Use of the presented method allows clinicians to detect subtle pathological changes from one image to the next, thus improving the quality of patient management in the ICU.

1 INTRODUCTION

In the ICU, clinical evaluation can rely heavily on diagnostic images such as portable chest radiographic images. The successive diagnostic images taken by a portable computed radiography (CR) system are helpful for indicating significant pathological changes of the patient, such as a collapsed lung or and improper tube placement within the patient.

However, image differences owing to different exposure settings, or patient and apparatus positioning, limit the accuracy of image comparison in the ICU, even for those images obtained from the same patient over a short treatment interval. Obviously it constrains the ability of the clinician to subtle changes that can be highly significant. An important problem is allocating the output dynamic ranges to display the clinically important part of the input code values. The process of selecting the relevant sub-range of input code values and constructing the proper mapping function from the input code values to the output display media is termed a tone-scale adjustment. Using a tone-scale method in CR images provides an optimal rendering result (Lee and Barski, 1997). There are also other methods (Barski and Metter, 1998) that provide improvements in contract enhancement for diagnostic imaging.

However, these methods do not address the problem of consistent rendering between images of the same patient taken at different times. Application of such tone-scale techniques is not likely to provide consistent rendering results, which makes accurate changes assessment by the ICU clinician difficult.

In this paper, we present a region of interest (ROI)-based lookup table (LUT) mapping method for diagnostic images that provides a consistent rendering result for images taken of the same patient at different times. This will help the clinicians compare images and track patient progress. First is a background segmentation step when the background of all the images (that may have different amounts of background content or no background content) are segmented. In the ROI selection step, the ROI region is located. These are the images of the tissue parts that are critical for clinicians to make a correct diagnosis. Next, an LUT constructed for the pixel values in the ROI. Then a toe-shoulder construction step is taken, constructing a LUT for very dark and very light regions. In the LUT mapping step, the pixel values in the input images are mapped to the corresponding pixel values in the output image.

The structure of this presentation is organized as follows: in section2, we introduce why and how the ROI is selected. In section3, the ROI-based LUT construction method is presented. Section4 reports the performance comparison result of the current method and of baseline method. Finally, the conclusion is drawn in Section 5.

2 AUTOMATED ROI SELECTION

After doing a background segmentation based on ICU's image histogram and difference histogram (Kuhn, 1999), we get an appropriate threshold for removing the background. A region-labeling operation can be done to prevent over-segmentation. Then we perform the automated ROI selection.

In ICU images, the position of the parts necessary for the clinicians' diagnosis varies. In some cased, they will only take up a little part of the image. The basic principle of automated ROI selection is to identify the RIO in each image automatically and adjust the image contrast values within the ROI to a suitable range for each image, so that comparison of one image to another is feasible.

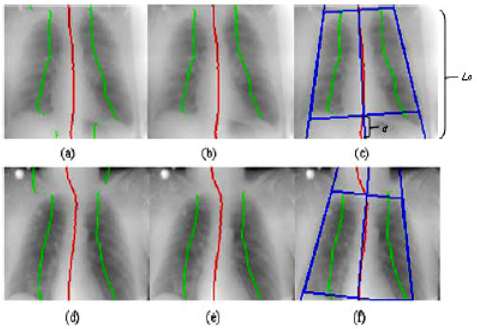


Figure 1: Automated region of interest selection; this is an example of selecting similar regions of interest for two images of the same patient.

ROI identification located key features (lung line, spine line) in an image and allows the correlation of two or more images accordingly. Figures1(c) and (f) show two chest X-ray images of the same patient with two automated regions of interest (ROI) selected.

First we use a median filter to resize the image, then a Gaussian filter for noise removal. Next, the locations of the spine line and lung line are detected (Amit and Mark, 2005). Fig.1 (a) and (d) show the spine and lung line detection. We search for the highest/lowest mean column value row by row. Connecting these points, we validate the lung line step (Fig.1 (b) and (e)), and combine and validate similar lung line parts based on gray-level and position.

With the approximate lung line and spine line determined, a spine-line-fitting step can be executed. This is performed by doing an iterative of the spine-line-fitting step. We search all the rows between the top and bottom of the lung lines. We then choose the fitting result that has the lower mean residual form these two. We then can get a trapezoid ROI for all the images of the same patient based on the spine line and the distance of the spine line to the lung line.

3 ROI-BASED LUT CONSTRUCTION

Once one or more ROIs have been identified, we can do the ROI-based LUT construction step.

First we identify the primary area o the image from the histogram data that is related only to the ROI. Points lp and rp represent left and right points, respectively, of the histogram data that is from the main range (2.5%-95%) in the ROI. After that, for each image, left points lp1 and lp2, and right points rp1 and rp2, are obtained. The goal of next few steps is to remap left points lp1 and lp2, and right points rp1 and rp2, to the corresponding points A1 and A2, in order to form consistent images in the output images.

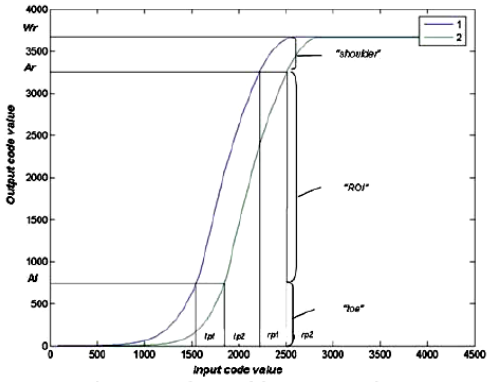


Figure 2: Lookup table construction.

Figure 2 shows how various portions of the image are remapped for consistent rendering. We can map the right point rp, obtained from the ROI of each input image, to the same value Ar in the output image that has been determined for the same patient. However, to accommodate the difference in patient position between two images of same patient, we proposed to use Ar for each image. Here, the

diaphragm in Fig.1 (a) is higher that that in Fig.1 (d). This difference can be best expressed by means of a proportion of distance d to column length Lc for each image as illustrated in Fig.1.

Given these considerations, Ar can be calculated using the following calculations to adjust the difference in patient position:

$$Ar_1 = p_1 \cdot \left(\frac{d}{L_c}\right) + p_2; Ar_2 = p_1 \cdot \left(\frac{d}{L_c}\right) + p_2; \\ Ar = \left(Ar_1 \cdot t\right) + Ar_2 \cdot (1-t) 0 \le t \le 1 \quad (1)$$

Where d and column length L_c are illustrated in Figure 1. p_1, p_2 and t are empirical parameters.

In out method, features used to determine the value Al include the difference $rp - lp$, and the value of $rp - lp/(spinedifference)$:

$$x = \overline{rp - lp}; Al' = ax^2 + bx + c \\ pdark = \frac{\overline{rp - lp}}{spdv - spuv} \\ Al = Al' + \max\left(a_1, \min\left(a_2 \cdot (pdark - a_0), a_4\right)\right) \\ Al = \max\left(\min Al, \min\left(\max Al, Al\right)\right) \quad (2)$$

$a, b, c, a_1, a_2, a_0, a_4, \min Al, \max Al$ are empirical parameters and $spdv, spuv$ (spine down-part value and spine up-part value) are the main gray-level range in the spinal region (10%-80%), which can be detected automatically. Note that the A1 can be justified differently by the ratio of $pdark$ for each image. Here we choose the same Al for all the images from the same patient.

After we get lp, rp, Al, Ar for each image, the LUT construction between lp and rp to Al and Ar can be applied. The mapping from $[lp, rp]$ to $[Al, Ar]$ is established based on the active rate (Lee, 2004) calculated in equation (3).

$$Activity[k] = \frac{\sum_{(i,j)\in ROI, img(i,j)=k} \sum_{u=i-3}^{i+3} \sum_{v=j-3}^{j+3} H(u,v,i,j)}{h(k)} \\ ActNor[i] = \frac{\ln\left(\frac{\sum_{i=-t}^{t} Activity[i+t] \cdot h[i+t]}{\sum_{i=lp}^{rp-1} h[i]} + 1\right) + 1}{\sum_{i=lp}^{rp-1} \left(\ln\left(\frac{\sum_{i=-t}^{t} Act[i+t] \cdot h[i+t]}{\sum_{i=lp}^{rp-1} h[i]} + 1\right) + 1\right)} \quad (3) \\ when h1 \le \left|img(i,j) - img(u,v)\right| \le hr, H(u,v,i,j) = 1$$

$activity[k]$ is the activity of intensity k and h(i) is the number of the pixel at that intensity.

Figure 3 shows an ICU image's active rate and example of LUT construction using equation(4) considering the active rate.

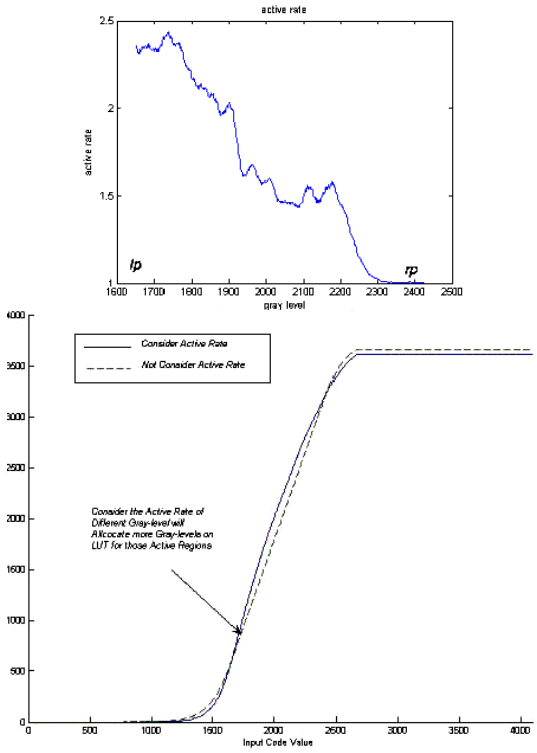


Figure 3: Active-rate in LUT construction.

$$LUT(lp) = Al \\ LUT(i+1) = LUT(i) + ActNor(i)(Ar - Al) ratio \\ + \left(\frac{(Ar - Al)(1 - ratio)}{rp - lp}\right) \quad (4) \\ Here, 0 \le ratio \le 1, when ratio = 0$$

In addition to mapping the ROI of the image, for darker or brighter regions, a toe-shoulder LUT construction step was performed for additional mapping, such as the toe region and the shoulder region in the LUT curve in Fig.3. The toe region was constructed for mapping the dark area in the image and the shoulder region was constructed for the bright area in the image.

4 PERFORMANCE

We collected 83 portable X-ray images from 19 patients. There were two to nine images of each patient. An experienced chest radiologist reviewed all the images from the 19 patients and provided a diagnosis that included the types of diseases detected

and any change in a patient's condition (improved/worsened). We compare the presented method with a baseline image enhancement technique that is an image optimization technique based on single image (Barski and Metter, 1998). An evaluation of the images from the 19 patients was performed in order to compare the overall consistency in the image and the lung areas as well as the ability to detect changes in patients' conditions against the radiologist's diagnosis.

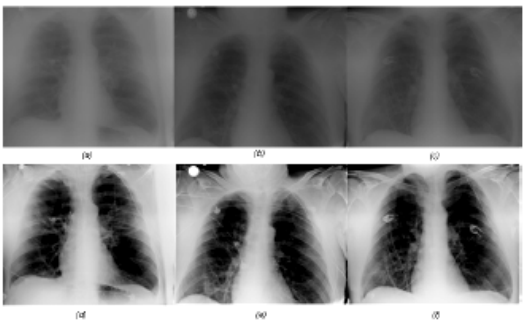


Figure 4: Processed image of the same patient. (a), (b), and (c) are raw images of the same patient, and (d), (e) and (f) are the processed result using the current method.

Figure 4 shows the processing result examples of a patient's ICU chest X-ray images. In the evaluation, all of the processed images form each patient were presented to a radiologist in the order of the processed images form the baseline method first and next the processed images from the presented method. The radiologist gave a rating on a 5-point acceptability scale where 1 is not acceptable and 5 is outstanding in terms of the consistency rendering effect demonstrated among the images presented for diagnostic purposes. Table 1 is the evaluation result. A t-test is also done to compare the baseline and current methods.

5 ROI-BASED LUT CONSTRUCTION

Our image-rendering technique reduces the variability in the image appearance caused by the differences in patient or apparatus positioning and image acquisition parameters. The improved consistency over the baseline image enhancement technique can potentially improve the overall workflow and patient management.

Thus, it is a method for enhancing diagnostic images taken at different time in order to provide consistent rendering for regions of interest.

Table 1: The evaluation result.

	Patient Number	Score					Average	T-test
		5	4	3	2	1		
Current Method	19	9	8	1	1	0	4.32	Accept P-Value =0.039
Baseline Method	19	2	9	5	3	0	3.52	
	Image Number	Score					Average	T-test
		5	4	3	2	1		
Current Method	83	42	30	9	2	0	4.35	Accept P-Value =0.0034
Baseline Method	83	27	34	17	4	1	3.99	

ACKNOWLEDGEMENTS

This paper is supported by Innovation Program of Shanghai Municipal Education Comission.

REFERENCES

Amit. Singhal, Mark Bolin, Hui Luo, 2005. "Inducing node specification in active shape models for accurate lung-field segmentation." Proc. SPIE, Vol. 5747, pp. 431-442.

G. Kuhn, 1999. "Method and apparatus for automatically location region of interest in a radiograph." U.S. Patent. 5896463.

H. C. Lee, L. Barski, R. Senn, 1997. "Automatic tone scale adjustment using image activity measure." U. S. Patent. 5633511.

H. C. Lee, 2004. Tone scale processing based on image modulation activity. U. S. Patent 6717698.

L. Barshi. R, Van. Metter, 1998. "New automatic tone scale method for computed radiography." Proc. SPIE Vol. 3335, pp. 164-178.

R. Van. Metter, D. Foos, 1999. "Enhanced latitude for digital projection radiography." Pro. SPIE, Vol. 3658, pp. 468-483.

INVESTIGATION OF ICA ALGORITHMS FOR FEATURE EXTRACTION OF EEG SIGNALS IN DISCRIMINATION OF ALZHEIMER DISEASE

Jordi Solé-Casals
Signal Processing Group, University of Vic, Sagrada Família 7, 08500 Vic, Spain
jordi.sole@uvic.cat

François Vialatte, Zhe Chen, Andrzej Cichocki
RIKEN Brain Science Institute, LABSP, 2-1 Hirosawa, Saitama, 351-0106 Wako-Shi, Japan
{*fvialatte, zhechen, cia*}*@brain.riken.jp*

Keywords: EEG, Alzheimer disease, ICA, BSS, Feature extraction.

Abstract: In this paper we present a quantitative comparisons of different independent component analysis (ICA) algorithms in order to investigate their potential use in preprocessing (such as noise reduction and feature extraction) the electroencephalogram (EEG) data for early detection of Alzhemier disease (AD) or discrimination between AD (or mild cognitive impairment, MCI) and age-match control subjects.

1 INTRODUCTION

Independent component analysis (ICA) is a method for recovering underlying signals from linear mixtures of those signals. ICA draws upon higher-order signal statistics to determine a set of ”components” which are maximally independent of each other.

The aims of this paper is to investigate which ICA algorithm is best adapted to deal as a preprocessing stage with EEG signals. In order to do that, we made different experiments with EEG data from Alzheimer and age-match control subjects. The evaluation was calculated in terms of measure of receiver operating characteristic (ROC) score.

The paper is organized as follows: in Section 2 we present experimental data characteristics used in the experiments. Section 3 is devoted to procedure and ICA algorithms used. In Section 4 we explain the measure that we will use for obtaining the experimental results, that are presented in Section 5. Finally, conclusions are presented in Section 6.

2 EXPERIMENTAL DATA

In the course of a clinical study, mutlichannel EEG recordings (Deltamed EEG machine) were recorded from 33 elderly patients affected by Alzheimer's disease and followed clinically (labeled AD set) and from 39 age-matched controls (labeled Control set), with electrodes located on 19 sites according to the 10-20 international system. This database was recorded in normal routine. Reference electrodes were placed between Fz and Cz, and between Cz and Pz. The sampling frequency was 256 Hz, with band-pass filter 0.17-100 Hz. Three periods of 5 seconds were selected in a ”rest eyes-closed” condition for each patients. In selecting these three independent sessions, an artifact rejection procedure was used to help minimize the artifact effect.

3 ICA AND BSS

3.1 Procedure

At the first stage, we apply principal component analysis (PCA) to perform dimensionality reduction. At the second stage, an ICA algorithm is implemented to perform BSS. The estimated output signal $\mathbf{y}_t$ are assumed to be the source signals of interest up certain scaling and permutation ambiguity.

In addition, if we are only interested in denoising or getting rid of specific component, we can set that specific output signal (say y_i) to zero while keeping other components intact, and apply back projection procedure to recover the original scene. In our experiments, in ranking the output components, we al-

ways select the one that has the least absolute kurtosis value (i.e., the one close to Gaussian by assuming zero kurtosis statistic for Gaussian signal, positive kurtosis statistic for super-Gaussian signal, and negative kurtosis for sub-Gaussian signal).

3.2 Selection of Candidate Algorithms

For comparison, we have selected seven representative ICA algorithms. The selection criteria of them are based on several factors: (i) computationally efficiency; (ii) robustness; (iii) fewer degree of freedom (such as the choices of learning rate parameter, nonlinearity, or number of iterations); (iv) preference to batch method.

Specifically, the following seven ICA/BSS algorithms are among some of most popular BSS methods in the literature: AMUSE, SOBI, JADE, Pearson-ICA, Thin-ICA, CCA-BSS and TFD-BSS.

The detailed description of algorithms are neglected here; for relevant references, see (Cichocki and Amari, 2002). All of algorithms are implemented in MATLAB, some of them are available for download from the original contributors (Cichocki et al.,).

For each algorithm, we have varied the number of independent components (namely, n), from 3 to 10, to extract the resultant uncorrelated or independent components.

4 PERFORMANCE EVALUATION

In signal detection/classification theory, a receiver operating characteristic (ROC) is a graphical plot of the sensitivity vs (1-specificity) for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented equivalently by plotting the fraction of true positives (TP) vs the fraction of true negatives (TN). Nowadays, the usage of ROC has become a common measure to evaluate the discrimination ability of the feature or classifier. Roughly, the discrimination ability or performance is measured by the area value underneath the ROC curve, the greater the value, the better is the performance (with 1 denoting perfect classification, and 0.5 denoting pure random guess).

Since the primary purpose here is to evaluate the features extracted from different ICA algorithms, we have focused on the comparison between ICA algorithms and the choice of number of independent components. In order to obtain the baseline, we choose two simple yet popular linear classifiers—the linear discriminant analysis (LDA) and linear perceptron. In calculating the ROC score, we have employed the leave-one-out (LOO) procedure.

The features we use to feed the linear classifier are the power values extracted from different frequency bands (θ, α, β, and γ). The ROC score is first calculated by using raw EEG data without any ICA preprocessing; this ROC score is regarded as a baseline for further comparison. For ICA feature extraction, we conduct the procedures of dimensionality reduction, source separation, component rejection, followed by backward projection. For each algorithm, we calculate their ROC score by varying the number of independent components from 3 to 10. Note that all the discrimination tasks are binary classification: AD against control subjects.

5 EXPERIMENTAL RESULTS

First, we calculated the ROC score for all ICA algorithm with varying number of independent components. All algorithms follow the similar-shape trend: compared to baseline, there is a positive gain in high-frequency bands using ICA; while for low-frequency bands, there is no need for using ICA because of their negative gains. In fact, the result is consistent with what was expected: since the SNR is poor in high-frequency bands, eliminating the independent component with the least absolute value of kurtosis would lead to a gain in SNR; consequently, the ROC score or its gain is greater.

Next, the comparison was conducted on three individual 5-second sessions' EEG recordings. By averaging these three independent data set, we also obtain the performance comparison. It can be seen from these results that for all independent data sets, the performance depends on the choice of the ICA algorithm as well as the choice of components. On the other hand, it is also obvious that by using ICA algorithms for feature extraction, it is possible to boost the ROC score performance (w.r.t. the baseline) around $\frac{0.7467-0.6193}{0.6193} = 20.6\%$ (data set 1), 15.6% (data set 2), and 10.2% (data set 3), assuming the best ICA algorithm (with optimum number of IC) is employed. This improvement is quite significant. The averaged ROC score against the number of independent components is plotted in Figure 1.

From Table 1, several noteworthy observations are in order:

- It seems that the optimum number of IC is 4, obtaining the highest mean ROC score (averaged over all ICA algorithms) 0.6536, followed by 0.6447 (IC=6). Overall, it seems the optimal range for the number of IC is between 4 to 7.

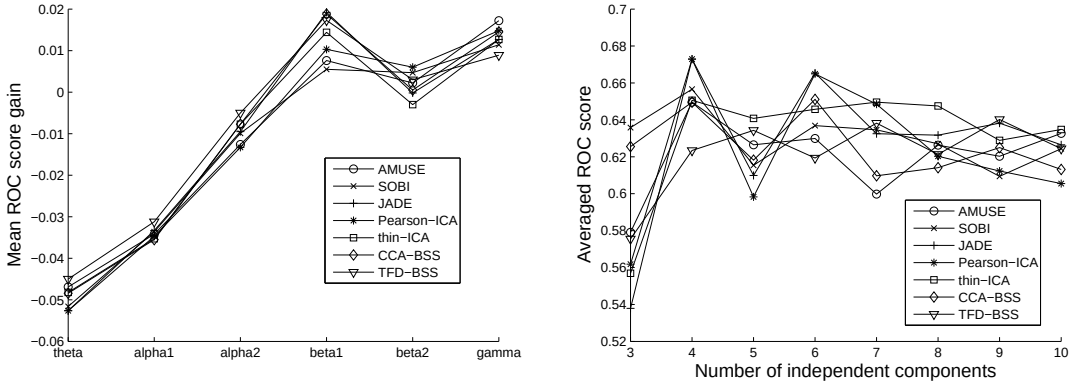


Figure 1: Left: The mean ROC score gain (averaged over 3 data bases and the number of independent components; with 0 as baseline) against frequency bands for 7 algorithms. Rigth: The averaged ROC score (over 3 data bases) comparison between different ICA algorithms with varying number of independent components.

- By averaging different numbers of IC, it seems the overall best ICA algorithms are Pearson-ICA and JADE (averaged from 4 to 7 components), or Thin-ICA and SOBI (averaged from 3 to 10 components).
- Overall, JADE and SOBI seem to give quite consistent performance for different number of components.

In addition, we can also compare the correct classification rate between different ICA algorithms with different setups. The results using LDA and linear perceptron classifiers are summarized in Table 2. Likewise, compared to the baseline correct classification rate, the performance with appropriate ICA algorithm give some more or less improvement.

6 CONCLUSIONS

In this work, we have proposed a measure or criteria to compare several popular ICA algorithms in the investigation of feature extraction of eeg signals in discrimination of Alzheimer disease. As a powerful signal processing tool used in the preprocessing step, ICA was found useful in artifact rejection, improving SNR, and noise reduction, all of which are important for the feature selection at the later stage. The ICA algorithms and the optimum choice of independent components are extensively investigated using simple linear classifiers and LOO procedure for calculating the resultant ROC scores and correct classification rate, both compared to their baselines.

It was found, in general, ICA algorithms are particularly useful for feature extraction in high frequency bands, especially on high alpha and beta ranges; where in low frequency bands, little gain has been obtained compared to the baselines. This is more or less anticipated, because EEG signals are usually contaminated by noise at high frequency bands, but are more resistant to noise at low frequency bands. Moreover, the optimum number of selected components seem to depend on the selected algorithms, but overall observations seem to indicate the number should be in the range from 4 to 7. Interestingly, this number is consistent with our early independent investigations (Vialatte and et al., 2005). In terms of overall average performance, it seem that the JADE, SOBI, thinICA, and CCABSS algorithms give more consistent and better results.

ACKNOWLEDGEMENTS

This work has been partly funded by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya and by the Ministerio de Educación y Ciencia under the grant TEC2007-61535/TCM

REFERENCES

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing*. Wiley, New York.

Cichocki, A., Amari, S., and et al. Icalab toolboxes. http://www.bsp.brain.riken.jp/ICALAB.

Vialatte, F. and et al. (2005). Blind source separation and sparse bump modelling of time frequency representation of eeg signals: New tools for early detection of alzheimer's disease. In *Proc. IEEE Work. Machine Learning for Signal Processing, pp. 27–32.*

Table 1: The ROC score comparison between ICA algorithms by averaging the results from three 5-second sessions. The baseline value (without ICA) of ROC score is 0.63. The bold fonts indicate the top two winners or the maximal two values in each column.

no. IC	AMUSE	SOBI	JADE	Pearson-ICA	Thin-ICA	CCA-BSS	TFBSS
3	0.5791	0.6358	0.5379	0.5618	0.5569	0.6255	0.5755
4	**0.6496**	**0.6566**	**0.6729**	**0.6729**	**0.6506**	**0.6496**	0.6234
5	0.6265	0.6154	0.6099	0.5983	0.6408	0.6182	**0.6343**
6	**0.6299**	**0.6369**	**0.6654**	**0.6649**	0.6457	**0.6511**	0.6193
7	0.5998	0.6346	0.6325	0.6485	**0.6496**	0.6097	**0.6382**
8	0.6263	0.6270	0.6317	0.6203	0.6475	0.6141	0.6213
9	0.6203	0.6094	0.6382	0.6123	0.6289	0.6250	0.6402
10	0.6327	0.6244	0.6265	0.6055	0.6348	0.6131	0.6244
	average from $\{2,3,4,5\}$ rows						
	0.6292	0.6210	**0.6498**	**0.6544**	0.6385	0.6094	0.6364
	average from all rows						
	0.6205	**0.6300**	0.6269	0.6231	**0.6318**	0.6258	0.6221

Table 2: Classification results using leave-one-out procedure. The bold font indicate the maximum value in each column.

no. IC	AMUSE	SOBI	JADE	Pearson-ICA	Thin-ICA	CCA-BSS	TFBSS
	LDA			baseline value (without ICA) 75%			
3	65.2778	65.2778	65.2778	63.8889	62.5000	61.1111	62.5000
4	69.4444	73.6111	70.8333	69.4444	68.0556	69.4444	61.1111
5	68.0556	66.6667	**77.7778**	69.4444	75.0000	73.6111	69.4444
6	72.2222	72.2222	76.3889	75.0000	70.8333	**77.7778**	69.4444
7	68.0556	73.6111	70.8333	**76.3889**	72.2222	72.2222	72.2222
8	73.6111	76.3889	72.2222	73.6111	72.2222	70.8333	70.8333
9	**76.3889**	73.6111	68.0556	73.6111	76.3889	69.4444	70.8333
10	70.8333	**75.0000**	73.6111	72.2222	**77.7778**	72.2222	**76.3889**
	linear perceptron			baseline value (without ICA) 62.5%			
3	59.7222	51.3889	54.1667	54.1667	45.8333	**68.0556**	54.1667
4	62.5000	**70.8333**	65.2778	**70.8333**	**68.0556**	56.9444	54.1667
5	59.7222	62.5000	56.9444	56.9444	68.0556	62.5000	62.5000
6	65.2778	62.5000	**70.8333**	**65.2778**	68.0556	65.2778	62.5000
7	59.7222	56.9444	65.2778	65.2778	62.5000	65.2778	**65.2778**
8	**65.2778**	62.5000	59.7222	59.7222	65.2778	65.2778	62.5000
9	62.5000	62.5000	**70.8333**	56.9444	68.0556	62.5000	62.5000
10	62.5000	62.5000	68.0556	59.7222	62.5000	65.2778	59.7222

USING WAVELET TRANSFORM FOR FEATURE EXTRACTION FROM EEG SIGNAL

Lenka Lhotska, Vaclav Gerla, Jiri Bukartyk
Gerstner Laboratory, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic
lhotska@fel.cvut.cz, gerlav@fel.cvut.cz

Vladimir Krajca, Svojmil Petranek
University Hospital Bulovka, Budínova 2, 18081 Prague 8, Czech Republic
krajca@fnb.cz

Keywords: EEG processing, wavelet transform, feature extraction.

Abstract: Manual evaluation of long-term EEG recordings is very tedious, time consuming, and subjective process. The aims of automated processing are on one side to ease the work of medical doctors and on the other side to make the evaluation more objective. This paper addresses the problem of computer-assisted sleep staging. It describes ongoing research in this area. The proposed solution comprises several consecutive steps, namely EEG signal pre-processing, feature extraction, feature normalization, and application of decision trees for classification. The work is focused on the feature extraction step that is regarded as the most important one in the classification process.

1 INTRODUCTION

The electroencephalogram (EEG), describing the electric activity of the brain, contains a lot of information about the state of patient health. It has the advantage of being non-invasive and applicable over longer time span (up to 24 hours if necessary). This is an important feature in case we want to follow disorders that are not permanently present but appear incidentally (e.g. epileptic seizure) or under certain conditions (various sleep disorders). Although the attempts to support EEG evaluation by automatic or semi-automatic processing have been made for a long time, there are still many problems to be solved. We try to contribute by our research to this effort. The main objective of the described work is the identification of the most informative features from sleep EEG records that could be used for automated (or semi-automated) sleep stage classification. Our approach to the analysis of human sleep uses wavelet transform (WT) and statistics for feature extraction and construction. The extracted and computed features are used as inputs for a decision tree (Quinlan, 1990) that is learned to classify individual sleep stages. We use for our experiment EEG sleep records rated by an expert, freely available and downloadable from the Internet (Kemp, 2007).

The paper is organized as follows. Section 2 describes sleep EEG signal and approaches to its evaluation. Methods used in our research are presented in Section 3. Section 4 is devoted to description of performed experiments. In Section 5 the results of experiments are discussed and the conclusion is presented in Section 6.

2 SLEEP AND ITS COMPUTER SUPPORTED CLASSIFICATION

Sleep is a non-uniform biological state that has been divided into several stages based on polysomnographic (PSG) measurements that include EEG, EMG, EOG, ECG, temperature, SpO_2 (oxygen saturation of the blood, recorded on the finger), respiration signals, as well as movement or body position. Polysomnography is usually performed over the duration of an entire night, or at least 6.5 hours, in order to investigate normal and disturbed sleep or vigilance (Bloch, 1997). Normal healthy sleep is organized into sequences of stages that typically cycle every 60 – 90 min. The most widely used standard for terminology and scoring of sleep

stages is the manual by Rechtschaffen and Kales (RK) (Rechtschaffen and Kales, 1968). A standard summary method is the hypnogram that graphically represents sleep stages in 20-30 second epochs. The PSG can be generally divided into epochs of 10, 20, 30, or 60 s, which are then visually classified into one of RK stages by a sleep technologist. The resulting time evolutionary description of sleep in terms of stages, termed hypnogram, is used by physicians for diagnosis. The Rechtschaffen and Kales manual details a complete process of recording and analysing sleep, which is followed by the vast majority of sleep laboratories, worldwide. On the basis of EEG (plus EOG and EMG), epochs can be scored into sleep stages:

- Stage 1 – shallow/drowsy sleep;
- Stage 2 – light sleep;
- Stage 3 – deepening sleep;
- Stage 4 – deepest sleep;
- Stage REM – dreaming sleep.

Stages 1 to 4 are frequently described as non-REM sleep, and stages 3 and 4 are described as slow wave sleep (SWS). Other scores are Wake (W) and Movement Time (MT). Since the depth of sleep changes continuously, the artificial demarcation of sleep stages by the RK classification is a simplification. The exact time of change of state is highly subjective and leaves room for interpretation by the physician who scores transitional epochs (e.g., Stage 1 and Stage 3) differently on different occasions (Schaltenbrand, 1996).

Studies have shown agreement between physicians performing scoring that ranges from 67% to 91% (Gaillard and Tissot, 1973), (Stanus et al., 1987), (Kim et al., 1992), depending on different scoring epoch lengths and number of readers. However it is necessary to remark that most data on interscorer agreement are based on the study of normal subjects. Processing of sleep recordings requires elaborate training and is time consuming and expensive. No generally accepted standard exists for automatic sleep staging, but computerization can improve efficiency and reduce cost (Doman, 1995), and enhance collaboration between laboratories (Kemp, 1993).

Various approaches to computer classification of PSGs have been used. Johnson et al. (Johnson et al., 1969) presented a spectral analysis study of the EEG in different stages, which was subsequently used by Larsen and Walter (Larsen and Walter, 1970) to develop an automated staging technique based on multiple-discriminant analysis. Agarwal and Gotman (Agarwal and Gotman, 2001) use a method based on the segmentation and self-organization technique. The following five steps are necessary to perform computer-assisted staging: segmentation; feature extraction; clustering; assignment of stages to different clusters of patterns; and optional smoothing of the hypnogram. The study (Agarwal and Gotman, 2001) shows that the greatest discrepancy occurs in Stage 1. The sensitivity and the specificity are 38.6% and 43.4%, respectively. This is to be expected in the highly transitional Stage 1. Stage 1 also has significant similarities to REM stage and can be considered as one stage away from Stage 1. Moreover, it is accepted that manual scoring of Stage 1 is the most subjective due to its transitional nature.

3 METHODS

In our study we have used similar procedure as Agarwal and Gotman and the same we used in one of our previous studies (Gerla, Lhotska, and Krajca, 2005). The sleep EEG signal classification comprises several steps: segmentation, feature extraction, feature normalization, feature selection, and generation of decision trees.

We have applied wavelet transform (Daubechies, 1992) to sleep EEG signal preprocessing. Mean of the signal is calculated and subtracted from a signal before WT is applied. Discrete Wavelet Transform (DWT) represented by a filter bank is employed for wavelet decomposition. Before the decomposition starts it is necessary to select a mother wavelet used for defining FIR filters and a level of a decomposition tree. For deciding which mother wavelet should be selected we consider the impulse response and amplitude frequency characteristics of the FIR filter specified by the corresponding mother wavelet. After the DWT is done we get approximation and detail coefficients as input data for further processing. Then the segmentation is performed.

Segmentation. The non-adaptive segmentation is employed. Non-adaptive or constant segmentation divides a signal into segments of a constant length. This kind of segmentation is basically the easiest one. The disadvantage of this method is that the segments are not necessarily stationary. The length of a segment is chosen regarding the character of data.

Feature extraction is the second most important part after wavelet decomposition. It is a process which changes representation of segments by extracting features from them. The aim is to select those features which carry most information about the segment. The statistic parameters are in principle very suitable for this purpose. We use autoregressive features and computed wavelet coefficients as well. We use the following parameters: average absolute

amplitude, maximal positive amplitude, maximal negative amplitude, maximal absolute amplitude, frequency weighted energy, sample mean, sample central moment, sample variance, statistical median, energy, and entropy. The autoregressive features are calculated from the transfer function of an *autoregressive model*, in which a present value x_n or future values x_{n+I} , i=1,2,… are estimated by using the previous values $\{x_{n-m}, \ldots x_{n-1}\}$ (Therrien, 1992). We can extract features from each source (an original signal, its first and second derivation) independently.

Feature normalization. Mean and standard deviation of extracted features are different. That could have a negative influence to the classification process, when a classifier uses distances between points in n-dimensional space. Before we start classification the features must be normalized to have the same mean and standard deviation. The features have normal distribution N(0,1).

Feature reduction. There are several different ways in which the dimension of a problem can be reduced. In this work Principal Component Analysis (PCA) (Smith, 2002) approach is used which defines new features (principal components or PCs) as mutually-orthogonal linear combinations of the original features.

Feature selection is considered successful if the dimensionality of the data is reduced and the accuracy of a learning algorithm improves or remains the same. Decision tree algorithms such as C4.5 can sometimes overfit training data, resulting in large trees. In many cases, removing irrelevant and redundant information can result in C4.5 producing smaller trees. The Chi-squared statistic is used for feature selection.

Classification. We have decided to use decision tree algorithms because they are robust, fast, and what is important especially in medical domain their results are very easy to interpret. In particular, the C4.5 algorithm has been applied, namely its J48 variant available in the Weka software tool (Frank et al., 2007).

Success rate of classification. As a measure of success rate we have used the overall accuracy of the classification. The overall accuracy is calculated as the relative number of correct decisions.

4 EXPERIMENTS

The main purpose of our experiments has been to find the most suitable wavelet decomposition and the most discriminative features to achieve good classification results. The analyzed EEG recordings are presented in the next section and then our experiments with EEG data are described.

4.1 Source of EEG Recordings

Our source of EEG recordings is The Sleep-EDF Database (Kemp, 2007). Four EEG recordings from different subjects were downloaded. The recordings were obtained from Caucasian males and females (21 - 35 years old) without any medication. They contain horizontal EOG, Fpz-Cz and Pz-Oz EEG, each sampled at 100 Hz. The recordings also contain the submental-EMG envelope, oro-nasal airflow, rectal body temperature and an event marker, all sampled at 1 Hz. Hypnograms are also added which are manually scored according to Rechtschaffen & Kales based on Fpz-Cz / Pz-Oz EEG instead of C4-A1 / C3-A2 EEG (Sweden et al., 1990).

Subjects, recordings and hypnogram scoring for the 4 sc* recordings are described in (Mourtazaev, 1995). Subjects and 4 st* recordings are more extensively described in (Kemp et al., 2000). The sleep stages Wake, Stage1, Stage2, Stage3, Stage4, REM and 'unscored' are coded in the file as binaries 0, 1, 2, 3, 4, 5, 6 and 9.

After reviewing the data we have found out that the classes in data are unevenly represented. Class 1 (Wake) is the most frequent one and class 5 (stage4) occurs sporadically. We have generated the training set in which all classes are equally represented. That means that a classification error caused by an unequal distribution of classes should be reduced.

4.2 Experiment 1

A goal of this experiment is to find features which contain the information about classes included in data. In other words the features should be highly correlated with the class. In our case we have six classes (wake, stage1, stage2, stage3, stage4, REM). This is a complex task and it is quite impossible to find only one feature to correlate with all classes.

We modify our goal to examine all features for every combination of two different classes and select the most significant feature for discriminating the classes from each other. There are 15 combinations so we get 15 features in total. We have chosen EEG sample (sc4002e0), which includes all 6 classes; each having 200000 samples (2000 seconds). For WT, the following setting has been used: level of decomposition tree 7; mother wavelet db6; wavelet coefficients used for feature extraction (2,1), (3,1), (4,1), (5,1), (6,1), (7,1), (7,0); segment length 10s.

The results of this experiment and the best features selected for classification of every combination of two different classes are shown in

Table 1: Results of experiment 1 and the best features selected for differentiation between couples of classes.

Stage	Wake	Stage 1	Stage 2	Stage 3	Stage 4	REM
class	1	2	3	4	5	6
1		96% - f1	97.5% - f2	99.5% - f3	99.5% - f5	98.9% - f1
2	96% - f1		*85% - f7*	91.5% - f8	98.5% - f9	*70% - f10*
3	97.5% - f2	*85% - f7*		*73% - f11*	94% - f12	*85% - f4*
4	99.5% - f3	91.5% - f8	*73% - f11*		*85% -f3*	94.5% - f13
5	99.5% - f5	98.5% - f9	94% - f12	*85% -f3*		99.4% - f6
6	98.9% - f1	*70% - f10*	*85% - f4*	94.5% - f13	99.4% - f6	

Table 2: Description of the used features.

feature	original name of a feature	source for extraction	wavelet coefficient	full name of the feature
f1	MeaAbV_1d_d2_Fpz-Cz	first derivation	D2 (2.1)	average absolute amplitude
f2	Energy_sg_d4_Pz-Oz	signal	D4 (4.1)	energy
f3	MeaAbV_Sg_d5_Pz-Oz	signal	D5 (5.1)	average absolute amplitude
f4	Energy_1d_d5_Fpz-Oz	first derivation	D5 (5.1)	energy
f5	Energy_1d_d5_Pz-Oz	first derivation	D5 (5.1)	energy
f6	FrWeiE_Sg_d6_Fpz-Cz	signal	D6 (6.1)	frequency weighted energy
f7	FrWeiE_1d_d5_Pz-Oz	first derivation	D5 (5.1)	frequency weighted energy
f8	FrWeiE_Sg_d5_Pz-Oz	signal	D5 (5.1)	frequency weighted energy
f9	MeaAbV_Sg_d5_Fpz-Cz	signal	D5 (5.1)	average absolute amplitude
f10	MeaAbV_Sg_d3_Pz-Oz	signal	D3 (3.1)	average absolute amplitude
f11	FrWeiE_Sg_d7_Pz-Oz	signal	D7 (7.1)	frequency weighted energy
f12	Energy_Sg_d6_Pz-Oz	signal	D6 (6.1)	energy
f13	Energy_1d_d6_Fpz-Cz	first derivation	D6 (6.1)	energy

Table 1. The names and sources of these features are presented in Table 2. There are five classification results below 90% as it is shown in Table 1. It means that we are not able to extract any single feature which can separate these particular combinations of two classes. There must be used a combination of features. We can see that there are two features (f1, f3) occurring not only once as most discriminative. Each of them is chosen to be the discriminative feature for two combinations. A set of features is therefore reduced and we have 13 features. Unfortunately 5 of the features (marked in italics in Table 1) are not good enough for classification and thus we have decided to perform another experiment where different wavelet coefficients are decomposed and other features are examined.

4.3 Experiment 2

The goal is implicated by the result of the previous experiment. There are five combinations of two classes (4x3, 6x2, 3x2, 5x4, 6x3) which are classified with success rate lower than 90% by using features extracted from the wavelet coefficients. Now we try to achieve more accurate results by extracting features from such wavelet coefficient that have the same frequency resolution. Frequency resolution of a wavelet coefficient depends on sample frequency of data (100Hz) and on the level of the wavelet coefficient. We may be able to find more specific features carrying more information about separability of classes. Two different settings and wavelet decomposition trees are used: 1. level of decomposition tree 4; mother wavelet db15; wavelet coefficients used for feature extraction (4,0), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (4,7); segment length 10s; 2. level of decomposition tree 5; mother wavelet db20; wavelet coefficients used for feature extraction (5,0), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (5,7), (5,8), (5,9), (5,10), (5,11), (5,12), (5,13), (5,14), (5,15); segment length 10s. Wavelet coefficients from the highest level of decomposition trees are used for feature extraction. They have the highest frequency resolution compared with others in the wavelet decomposition tree. We can assume that the features extracted from these coefficients carry different piece of information about classes.

The results of this experiment have not been so successful as we have expected. The only feature that has brought relatively significant improvement of differentiation between two classes (by 7.5%) has been average absolute amplitude FPz-Cz (wavelet

coefficient (5,1)). The experiment has shown that application of other mother wavelets (db20 or higher) and different wavelet decomposition trees could result in finding new more discriminative features.

4.4 Experiment 3

The final experiment has been divided into three parts, namely using different groups of classes. EEG recordings (sc4012e0, st7022j0, sc4102e0) have been used as testing sets for this purpose.

Part 1. We have classified data into all six classes using all features as described in experiment 1. The results have verified our assumption that the features f3, f4, f7, f10 and f11 which do not separate classes well (see Table 1) decrease final classification accuracy.

Part 2. Based on the experiment 1 we have tried to distinguish among four classes, organized in two groups, namely (1, 3, 5, 6) and (1, 4, 5, 6). We have used six features from the original 15 for each group. The classification results for the first group have been negatively influenced by the feature f3 and for the second group by the feature f4.

Part 3. We have verified well discriminating features discovered in experiment 1. For this purpose we have selected three groups of three classes each that can be separated very well by these features. The three groups are composed of the following classes (1, 5, 6), (1, 2, 5) and (1, 4, 6).

All results are summarized in Table 3. The record sc4102e0 has not been used in those experiments where class 5 has been tested because it does not contain any segment belonging to class 5. The classification results are as we have assumed. They are mainly affected by low discriminability between classes 2 (stage1) and 3 (stage2) and classes 2 (stage1) and 6 (REM).

Table 3: Results of experiment 3 (success rate of classification).

classes	sc4002e0	sc4012e0	st7121j0	sc4102e0
1,2,3,4,5,6	72.1%	69.5%	63%	x
1,3,5,6	87.2%	87.6%	78.4%	x
1,4,5,6	87.8%	82.9%	73.2%	x
1,5,6	98.3%	94.5%	92%	x
1,2,5	97.2	92%	87%	x
1,4,6	96.5%	91.3%	90%	87%

5 DISCUSSION

Tables 1 and 2 are used for the interpretation. When we look at Table 2 we can see that all features extracted for the classification task in the experiment 1 are based on energy, mean absolute amplitude and frequency weighted energy. These features reflect the changes of energy in the given wavelet coefficient which is related to a specific frequency spectrum. This is very important as we see later. Now we try to explain why we have got the results of classification shown in Table 1. When we look at this table we can see that successful classifications are for the classes classified with features extracted from wavelet coefficients which have the frequency spectrum same as the frequency spectrum only a single class in the set of two classes has. That means that the feature used for such classification has high energy for this class and small energy for the other one. Then we can simply use a threshold to separate these two classes from each other. When we look at Table 1 again we can see that all successful classification results (success rate higher than 90%) are achieved between classes with mutual distance more than one class, for example, classes 1x3, 1x4, 2x5 etc. It is because the distance between these classes is quite long which is required for successful classification. An exception is the class 1 which is classified correctly in all cases. When we examine frequency spectra of classes 1 (Wake) and 2 (stage1), we find out that they are well separable. However we have to note that there exists overlapping (some frequencies occur in neighbouring stages). Therefore poorer classification result (below 90%) is for classes just next to each other (2x3, 3x4, 4x5 and 6x2). Unfortunately we have not yet found any feature better describing the classes by using different wavelet decomposition (experiment 2). The results of classification in experiment 3 are affected by this fact as well. In the following paragraph we suggest some ideas which could improve classification of sleep EEG data.

The frequency resolution of wavelet coefficients in level 5 of a wavelet decomposition tree is 3.12Hz. This decomposition is used in experiment 2. It was not detailed enough for distinguishing incorrectly classified classes. So we propose to make the frequency resolution higher by getting wavelet coefficients from level 6 (frequency resolution 1.57Hz) or even level 7 (0.78Hz). For these purposes we must ensure that the filter used for decomposition has steep frequency characteristic. We would recommend to use mother wavelets db20 and higher. If this condition is satisfied the results would not be influenced by leakage of other frequency components (antialiasing).

6 CONCLUSIONS AND FUTURE WORK

Sleep problems belong to the most common serious neurological disorders. Reliable and robust detection of these disorders would improve the quality of life of many people. The implemented methods allow automatic classification of EEG signals. The approach has been tested on real sleep EEG recording for which the classification has been known. We have focused on discovering the most significant features which would be highly correlated with classes of data. Our experiments have been based on the selection of a single feature to separate data belonging to two classes. There have been many other features with good selection results. The most frequent ones have been autoregressive features representing the order of used AR model and error of AR model. We have determined some features and wavelet coefficients which are best suited for classification of sleep EEG data. The future work will be focused on exploitation of other types of mother wavelets, using higher level of wavelet coefficients as source of features, and more sophisticated classifiers.

ACKNOWLEDGEMENTS

This work has been supported by the research program “Information Society” under Grant No. 1ET101210512 “Intelligent methods for evaluation of long-term EEG recordings“.

REFERENCES

Agarwal, R., Gotman, J., 2001. Computer-assisted sleep staging, *IEEE Trans. on Biomed. Engineering*, 48 (12), pp. 1412-1423.

Bloch, K.E., 1997. Polysomnography: a systematic review, *Technology and Health Care*, 5, pp. 285-305.

Daubechies, I., 1992. *Ten lectures on Wavelets*, CBMS-NSF, SIAM, 61, Philadelphia, Pennsylvania, USA.

Doman, J., Detka, C., Hoffman, T., Kesicki, D., Monahan, J.P., Buysse, D.J., Reynolds III, C.F., Coble, P.A., Matzzie, J., Kupfer, D.J., 1995. Automating the sleep laboratory: implementation and validation of digital recording and analysis, *Int. J. Biomed. Comput*., 38, pp. 277-290.

Frank, E., Hall, M., Trigg, L. Weka – Data Mining Software in Java. Internet site address: http:// www.cs.waikato.ac.nz.

Gaillard, J.M., Tissot, R., 1973. Principles of automatic analysis of sleep records with a hybrid system, *Comput. Biomed. Res*., 6, pp. 1-13.

Gerla, V., Lhotská, L., Krajča, V., 2005. Utilization of Time Dependence in EEG Signal Classification. *The 3rd European Medical and Biological Engineering Conference EMBEC´05 Prague.*

Johnson, L., Lubin, A., Naitoh, P., Nute, C., Austin, M., 1969. Spectral analysis of the EEG of dominant and nondominant alpha subjects during waking and sleeping, *Electroencephalogr. Clin. Neurophysiol.*, 26, pp. 361-370.

Kemp, B., 1993. A proposal for computer-based sleep/wake analysis, *J. Sleep Res*., 2, pp. 179-185.

Kemp, B., Zwinderman, A., Tuk, B., Kamphuisen, H., Oberyé, J., 2000. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE-BME* 47(9), pp. 1185-1194.

Kemp, B., 2007. Sleep Recordings and Hypnograms in European Data Format. The Netherlands. Internet site address: http://www.physionet.org/ physiobank/ database/sleep-edf.

Kim, Y., Kurachi, M., Horita, M., Matsuura, K., Kamikawa, Y., 1992. Agreement in visual scoring of sleep stages among laboratories in Japan, *J. Sleep Res*., pp. 58-60.

Larsen, L.E., Walter, D.O., 1970. On automatic methods of sleep staging by EEG spectra, *Electroencephalogr. Clin. Neurophysiol*., 28, pp. 459-467.

Mourtazaev, M., Kemp, B., Zwinderman, A., Kamphuisen, H., 1995. Age and gender affect different characteristics of slow waves in the sleep EEG. *Sleep* 18(7), pp. 557-564.

Quinlan, J.R., 1990. Decision trees and decision making. *IEEE Trans System, Man and Cybernetics*, 20(2), pp. 339-346.

Rechtschaffen, A., Kales, A. (eds.), 1968. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects, Brain Inform. Service/Brain Res. Inst., Univ. California, Los Angeles.

Schaltenbrand, N. et al., 1996. Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients, *Sleep*, 19, pp. 26-35.

Smith, L.I., 2002. *A tutorial on Principal Components Analysis*, University of Otago, New Zealand, 2002.

Stanus, E., Lacroix, B., Kerkhofs, M., Mendlewicz, J., 1987. Automated sleep scoring: A comparative reliability study of algorithms, *Electroencephalogr. Clin. Neurophysiol*., 66, pp. 448-456.

Sweden, B., Kemp, B., Kamphuisen, H., Velde, E., 1990. Alternative electrode placement in (automatic) sleep scoring (Fpz-Cz / Pz-Oz versus C4-A1 / C3-A2). *Sleep* 13(3), pp.279-283.

Therrien, C.W., 1992. *Discrete Random Signals and Statistical Signal Processing.* Englewood Cliffs (NJ): Prentice Hall.

PARAFAC CLASSIFICATION OF LAMB CARCASS SOFT TISSUES IN COMPUTER TOMOGRAPHY (CT) IMAGE STACKS

Jørgen Kongsro
Animalia – Norwegian Meat Research Centre, P.O. Box 396 Økern, N-0513 Oslo, Norway
jorgen.kongsro@animalia.no

Keywords: Lamb carcass, Tissue, PARAFAC, Multi-Way Analysis, Computer Tomography, Classification, Image, Stacks.

Abstract: Computer Tomography is shown to be an efficient and cost-effective tool for classification and segmentation of soft tissues in animal carcasses. By using 15 fixed anatomical sites based on vertebra columns, 120 lamb carcasses were CT scanned in Norway during autumn of 2005. Frequency distributions of CT values (HU [-200,200]) of soft tissues from each image were obtained. This yielded a 3-way data set (120 samples * 400 CT values * 15 anatomical sites). The classification of the soft tissues was done by multi way Parallel Factor Analysis (PARAFAC), which resulted in 3 components or soft tissues classified from the images; fat, marbled and lean muscle tissue.

1 INTRODUCTION

Computer Tomography is based on the attenuation of X-ray through a body. There is high correlation between the density of the body or body components, and the X-ray attenuation measured. This relationship is used to estimate the body composition, volume or weight of a biological sample. The attenuation of X-rays is visualized by reconstruction of 360^{o} rotation of X-ray tube in a CT tomogram or CT image. Image data from Computer Tomography can be orientated in different ways. Single slice tomograms can be handled like 2-way (rows*columns) data arrays. Stacks of tomograms from 3D samples are often orientated as multi-way data arrays (rows*columns*stack). Combining CT data with other types of data, like MRI etc., can also yield multi-modal data arrays which can be handled either in a multi-dimensional fashion or be unfolded prior to analysis. Unfolding of multi-way data may sometimes lead to poor estimation and interpretability of variation between the different stacks or batches in a multi-way data array.

There are two primary ways to perform classification. Supervised classification (1), where classes are known in advance (*a priori*), and un-supervised classification (2), where classes are not known in advance. For classification of soft tissues from CT images of lamb carcasses, it can be difficult to obtain solid *a priori* knowledge or reference data of classes. Traditionally, reference data has been collected by using destructive dissection. This procedure is both expensive and not very accurate due to differences operators / butchers (Nissen et al. 2006). The accuracy of classification of tissues may be influenced the accuracy of the reference method used, i.e. for calibration purposes or detection of false negatives or positives. By using non-supervised classification, validation techniques can ensure that the model works for new data and finds the optimal or true number of classes in the model. The non-supervised approach will not be affected by reference or *a priori* error.

Parallel Factor Analysis (PARAFAC) is one method designed to analyze and decompose multi-way data, and was introduced by Harshman in 1971 for Psychometrics. The PARAFAC method can be used as a non-supervised classification tool to classify soft tissues in CT image stacks sampled from whole lamb carcasses.

The main purpose of this study is to apply PARAFAC decomposition of multi-way CT image data array as a classification tool of different lamb carcass soft tissues.

2 MATERIALS & METHODS

2.1 Sampled Animals

120 lambs from a single Norwegian abattoir were sampled according to an experimental design from

August to September in 2005. The design was set up to cover the variation in all levels of fatness in the carcasses, and the principle of over-sampling at the extremes was applied (Engel et al. 2003): The carcasses were sampled in three groups; low, intermediate and high level of fatness. Selection was made using fatness score from the EUROP carcass grading system for lamb in Norway. Low fatness equals –2 standard deviations (st.dev.) and below mean value. High fatness + 2 std and above mean value (Kirton et al., 1995). Intermediate between high and low (table 1). 40% of the samples were selected for each of the groups low and high fatness and 20% selected for intermediate fatness (Tab. 1), yielding a 40-20-40 grouping of the designed samples. In addition, two subsets of equal size (50-50) were constructed for validation by split-half-analysis.

Table 1: Sampling and experimental design.

n = 120	Low		Mid		High	
	%	n	%	n	%	n
Design[1]	40	48	20	24	40	48
Subset 1[2]	38	23	20	12	42	25
Subset 2[2]	38	23	20	12	42	25

1 40-20-40 design for sampling

2 Data subsets for Split-half analysis

2.2 Computer Tomography

2.2.1 Settings

The lambs were scanned at the Norwegian University of Life Sciences using a Siemens Somaton Emotion CT Scanner. Two persons were involved in the scanning of lamb carcasses: one operation the scanner, and the other preparing and entering the carcasses into the machine. The capacity of this procedure was approximately 8-10 carcasses per hour. The protocol for CT scanning is described in Table 2.

Table 2: CT protocol used for scanning of lamb carcasses.

Topogram	Sequence
100 mA	170 mAs
130 kV	130 kV
Slice width: 2.0 mm	Scan time: 0.8s
Length: 1024 mm	Slice width: 3 mm
Tube position: AP	Number of scans: 15
Direction: Caudiocranial	Direction: Caudiocranial
Kernel: T80s (sharp)	Kernel: B50M
Window: 256-64	Window: 100-50
	Field of view (FOV): 400

2.2.2 Anatomical Sites

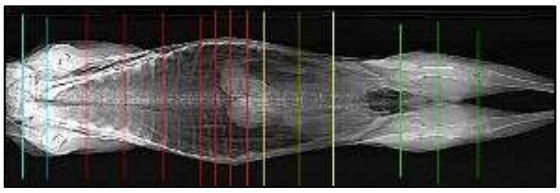


Figure 1: Scanning sites CT, lamb carcass.

Fifteen (15) anatomical scanning sites spanning the entire carcass were selected from a topogram using spine vertebras as fixing points (Fig. 1). A topogram is a survey picture produced by the CT-scanner. It is obtained by fixation of the X-ray tube in the upper position and moving the object at constant speed through the gantry. Each colour of the lines represents anatomical sections of the carcass (cervical, thoracic, lumbar, sacral and caudal). The anatomical sites were selected to span the entire variation of the carcass, but the number of images was limited due to capacity. High X-ray dose (170 mAs) was selected to increase the resolution of the tomograms. The anatomical sites collected from the mid-section of the carcass, were selected using literature reference sites for grading of lamb carcasses (Berg et al. 1997;Bruwer et al. 1987;Chandraratne et al. 2006;Chandraratne, Kulasiri, & Samarasinghe 2007;Cunha et al. 2004;Jones et al. 1992;Kirton et al. 1995). In addition to literature reference sites, additional sites on the leg and shoulder were added using spine vertebras as fixing points.

2.2.3 Import and Pre-processing of Images

The CT scanner generated images in DICOM format, which is a common medical image format. The images were imported into MATLAB using the Image Processing Toolbox routine dicomread.

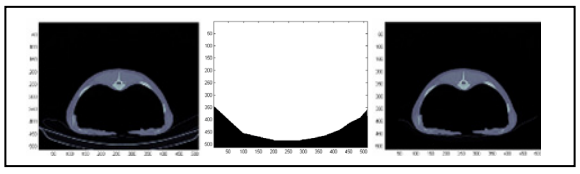


Figure 2: Raw image, Binary image for arithmetic extraction and processed image after extraction.

In the raw CT images, the couch material (non-carcass component) was visible (Fig. 2). This was removed using arithmetic extraction in MATLAB. The extraction was performed using image array

multiplication, subtracting the couch material from the raw image using a binary image (Fig. 2) containing zeros and ones to remove the couch area (zeros) of the image. The lamb area of the image was now extracted and ready for further analysis (fig.3).

2.2.4 Frequency Distribution of Pixel Values (HU)

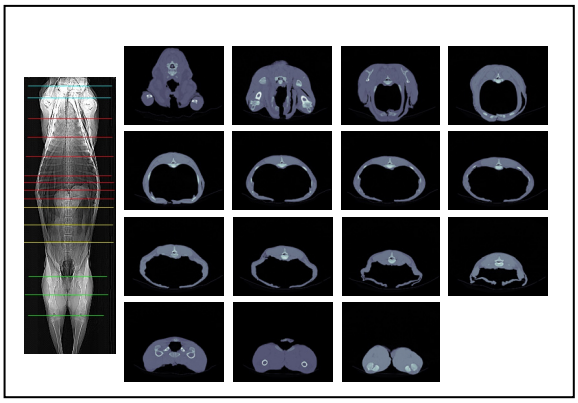


Figure 3: 15 pre-processed CT images from all scanning sites, from neck (1) to knee joint of leg (15).

A frequency distribution of the signal intensities (pixels) was generated for each anatomical site (Dobrowolski et al., 2004) (Fig. 5) from each of the pre-processed images (Fig. 5) using the frequency of Hounsfield Units [HU] in the interval [-200,200]. HU is related to density of biological tissues, where 0 is regarded as the HU of pure water. The interval of 400 HU is expected to cover the soft animal tissues (fat and muscle) in the CT images (Dobrowolski et al. 2004;Romvari et al. 2002). Each image was represented as a frequency distribution 2-way array [1 x 400]. For each sample, 15 images were generated, generating a 3-way array [1 x 400 x 15], giving a [120 x 400 x 15] data array for the entire samples.

2.3 PARAFAC

PARAFAC is a generalization of Principal Component Analysis (PCA) to higher order arrays (Bro 1997). Decomposition of the data array is made into triads or trilinear components (Fig. 4), but instead of one score vector and one loading vector as in bilinear PCA, each component consist of one score vector and two loading vectors (trilinear). PARAFAC is regarded as a "strong" multi-way method utilizing the multi-mode structure for modelling without unfolding, and providing other attractive features (Huang et al. 2003).

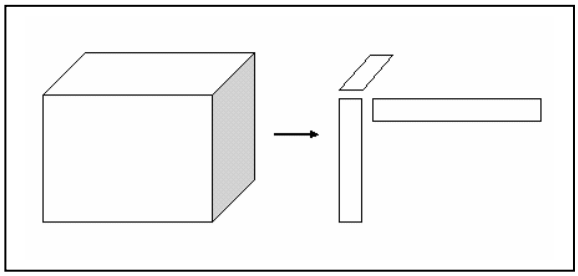


Figure 4: PARAFAC decomposition of a 3-way data array.

In this study, PARAFAC is used to decompose the landscapes of the frequency distributions and anatomical positions into a number of trilinear components (f).

$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} e_{ijk}$$

The element x_{ijk} represent the landscape of histogram spectra and anatomical positions of the lamb carcass sample i, frequency distribution j, anatomical position k. The landscapes are decomposed into sample scores a_{if}, frequency distribution loadings b_{jf} and anatomical position loadings c_{kf} for each factor f or PARAFAC component f. The residuals e_{ijk}, contains variation not explained by the model.

The PARAFAC components will be estimates of the CT histogram signals from the individual chemical components (fat & lean) if the data are approximately low-rank trilinear and when the correct number of components is used. If the optimal case is found, the scores for each of the components represent the relative content of carcass soft tissues. The number of components will represent the non-supervised classes of soft tissues suggested by the validated PARAFAC model.

PARAFAC models of CT image histogram landscapes were estimated with 1 to 4 components. The models were mean-centred since this has proven to yield the best result and interpretation of the figures. Since each component is expected to represent a single peaked frequency distribution (one local maxima) of a soft tissue, unimodality constraints was imposed on the model (Johansen et al. 2006). When calibrating PARAFAC models of CT images against a dissection reference, unimodality constraints seemed to yield the most accurate results (Johansen, Egelandsdal, Røe, Kvaal, & Aastveit 2006).

The correct number of components was selected using core consistency and split-half analysis (Bro 1997) as validation tools. The split was done using a 50-50 split of the actual designed (table 1) samples, estimating independent PARAFAC models for both split data subsets. Due to the uniqueness of the PARAFAC model, the same loadings will be obtained from different samples if the samples reflect the same CT histogram variables, when the correct number of components is chosen and enough data are available in each of the split data subsets (Andersen & Bro 2003).

All the models were constructed using the PLS_Toolbox 4.0, August 10, 2006, Copyright Eigenvector Research, Inc. 1995-2006 for MATLAB, the Image Processing Toolbox V5.3 (R2006b) for MATLAB and MATLAB 7.3.0.267 (R2006b), August 03, 2006 © 1984-2006The MathWorks Inc.

3 RESULTS & DISCUSSION

3.1 Landscapes

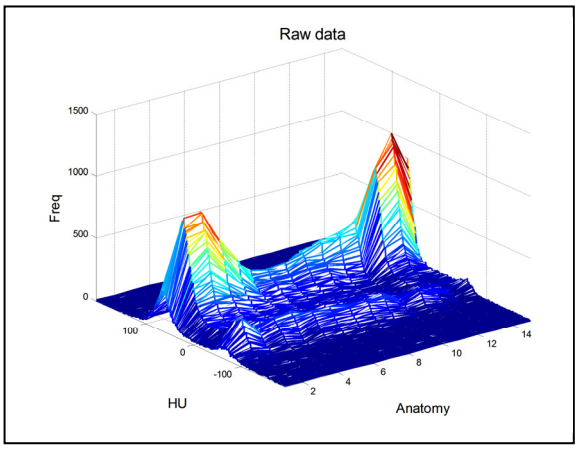


Figure 5: 3D CT histogram landscape of one sample, raw data.

From the landscape, a distinctive frequency distribution that appears between different anatomical sites is revealed (Fig. 5). There are two peaks identified as the shoulder site and the leg anatomical site. These sites are the "muscular" parts of the carcass (leg muscles), and therefore provide a high response or histogram intensity (number of pixels). There are two ridges in the landscapes, one larger than the other. The large ridge is identified as the lean tissue, and the small ridge as the fat tissue part. When comparing very fat animals with very lean animals, the fat ridge is almost absent in the very lean animals. These observations will be further investigated in the PARAFAC analysis.

3.2 PARAFAC

Table 3: PARAFAC diagnostics. Full model (n=120). # of components, explained variance, core consistency, number of iterations and computation time (s).

#	Expl. Var. (%)	Core cons.	# iter	Time (s)
1	50.06	100	5	6
2	66.36	95	11	12
3	73.18	78	18	20
4	76.80	0	23	28

The results from the PARAFAC models are shown in Table 3. Three components seem to yield a consistent model, with relatively low number of iterations and computation time.

In order to validate the appropriate number of components in the model, the results from the split-half analysis is shown in Figure 6. The figure shows the frequency distribution loadings for 1 to 4 components, were subset 1 has solid lines, and subset two dashed lines. Due to the uniqueness of the PARAFAC model, the same loadings should be obtained if the samples reflect the same CT histogram and anatomical site pattern when the optimal number of components is chosen. The solid and dashed lines seem to be correlated for the 1-, 2- and 3-component models, while for the 4-component model; the solid and dashed lines do not correlate. Thus, the model seems to be valid with 3 components.

The PARAFAC decomposition of the CT histogram landscapes is shown in Figure 7, were the raw landscape in Figure 5 is decomposed into three PARAFAC components. The 1st component seems to model the fat tissue in the frequency distribution, and the 2nd seem to model muscle tissue. The 3rd component seems to model very lean muscle tissue. Component 2 and 3 seem to be two types of muscle tissue, "marbled" or muscle tissue with higher fat content (# 2) and lean muscle tissue (# 3)

For practical purposes, the PARAFAC models yields a better understanding of the uniqueness and nature of the CT value (HU) frequency distribution. From the images in Figure 8, the PARAFAC loadings were applied to a test image from a lamb carcass belly. Loadings above a manually set baseline (0.02) were selected to ease interpretation of the test images. The first image from left represents the total area of soft tissue. The 2nd image

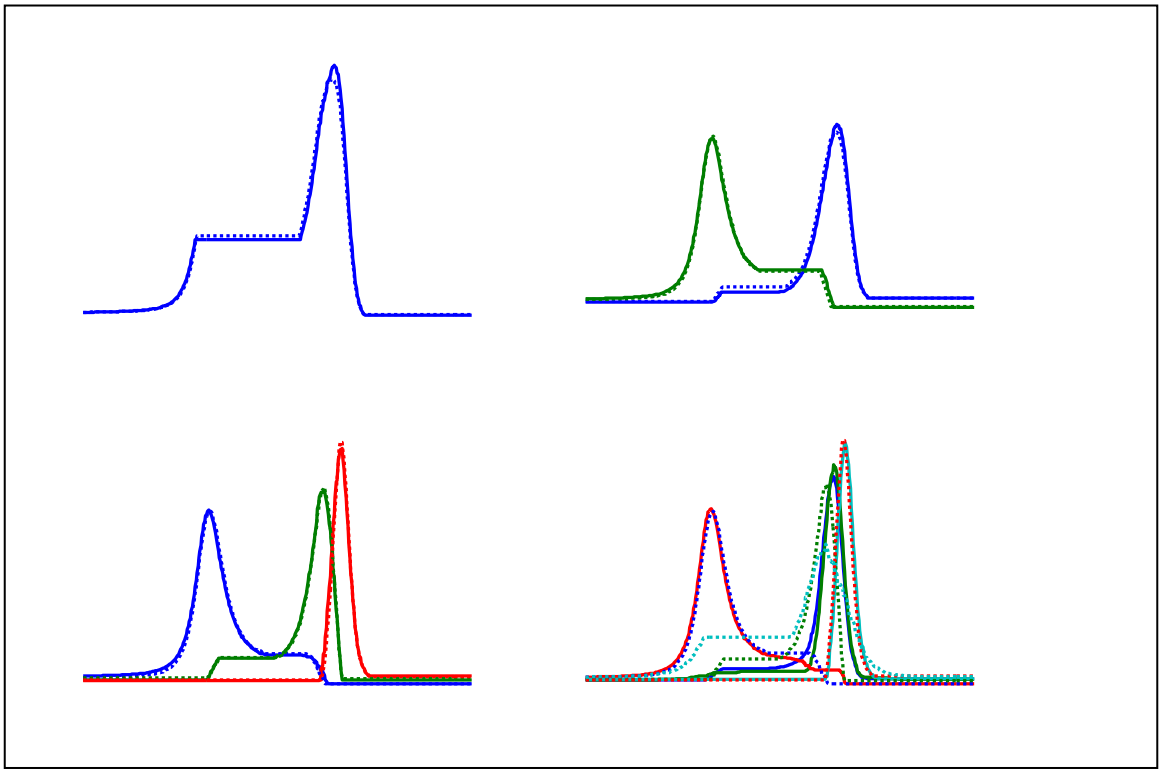


Figure 6: Validation of PARAFAC components (split-half analysis). 1 to 4 component PARAFAC models. 1 –and 2 – component model (top), 3- and 4- component model (bottom).

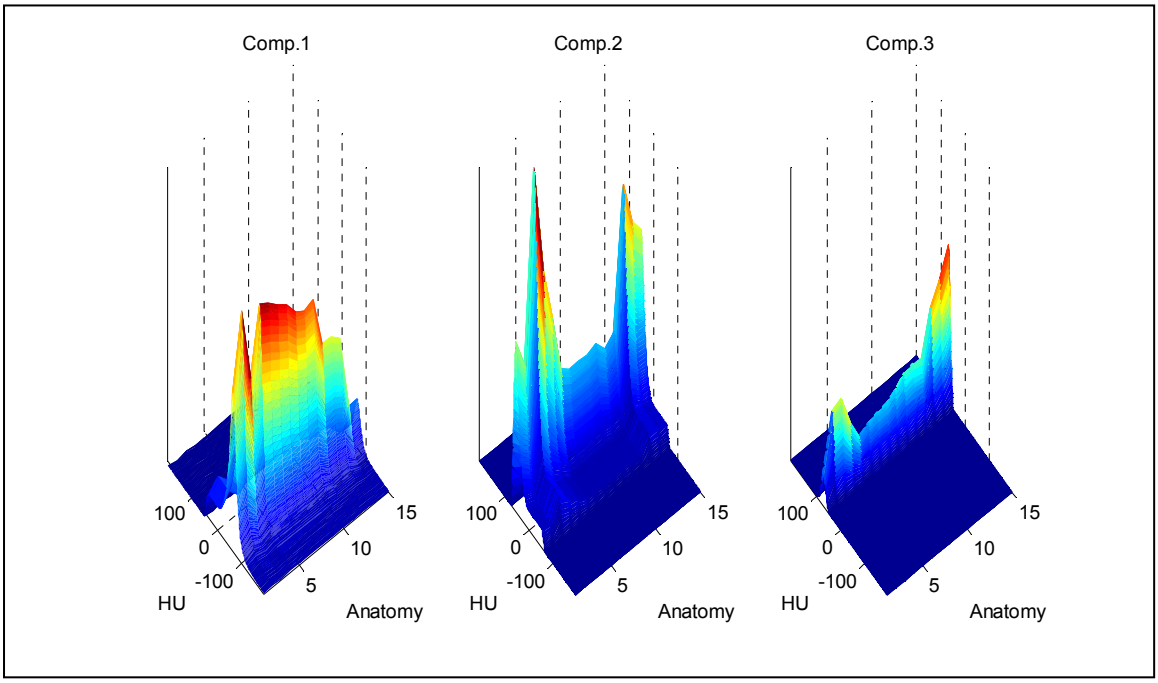


Figure 7: PARAFAC decomposition of a 3D CT frequency distribution. 3 components or classes identified. # 1 represent fat tissue, # 2 muscle tissue with marbling fat and the 3rd lean muscle tissue.

Figure 8; PARAFAC CT value frequency loadings > 0.02 applied on CT image from belly. From left, soft tissue HU= [-200, 200], # 1, # 2, # 3.

represents component 1, the 3rd component 2 and 4th (right) component 3. When inspecting the images visually, # 1 represent fat tissue, # 2 muscle tissue with marbling fat and the 3rd lean muscle tissue. PARAFAC yields a consistent decomposition of the 3D frequency distribution of the CT images, and selected 3 unique soft tissue components representing fat, and two types of muscle tissue.

4 CONCLUSIONS

This paper presents modelling and decomposition of multi-way array CT image data, using PARAFAC as a non-supervised classification tool for different lamb carcass soft tissues. Multi-way modelling applying PARAFAC did yield sensible interpretation of the 3D CT value frequency distribution. Three components or classes of soft tissues were extracted from the model; fat, marbled and lean muscle.

ACKNOWLEDGEMENTS

This study was sponsored by grant 162188 of the Research Council of Norway, as part of a Ph.D. study program. I am grateful for the assistance from Prof. Rasmus Bro at the Faculty of Life Sciences, University of Copenhagen.

REFERENCES

Andersen, C. M. & Bro, R. 2003, "Practical aspects of PARAFAC modeling of fluorescence excitation-emission data", *Journal of Chemometrics*, vol. 17, no. 4, pp. 200-215.

Berg, E. P., Neary, M. K., Forrest, J. C., Thomas, D. L., & Kauffman, R. G. 1997, "Evaluation of Electronic Technology to Assess Lamb Carcass Composition", *Journal of Animal Science*, vol. 75, no. 1997, pp. 2433-2444.

Bro, R. 1997, "PARAFAC. Tutorial and applications", *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149-171.

Bruwer, G. G., Naude, R. T., Dutoit, M. M., Cloete, A., & Vosloo, W. A. 1987, "An Evaluation of the Lamb and Mutton Carcass Grading System in the Republic of South-Africa .2. the Use of Fat Measurements As Predictors of Carcass Composition", *South African Journal of Animal Science-Suid-Afrikaanse Tydskrif Vir Veekunde*, vol. 17, no. 2, pp. 85-89.

Chandraratne, M. R., Kulasiri, D., Frampton, C., Samarasinghe, S., & Bickerstaffe, R. 2006, "Prediction of lamb carcass grades using features extracted from lamb chop images", *Journal of Food Engineering*, vol. 74, no. 1, pp. 116-124.

Chandraratne, M. R., Kulasiri, D., & Samarasinghe, S. 2007, "Classification of lamb carcass using machine vision: Comparison of statistical and neural network analyses", *Journal of Food Engineering*, vol. 82, no. 1, pp. 26-34.

Cunha, B. C. N., Belk, K. E., Scanga, J. A., LeValley, S. B., Tatum, J. D., & Smith, G. C. 2004, "Development and validation of equations utilizing lamb vision system output to predict lamb carcass fabrication yields", *Journal of Animal Science*, vol. 82, no. 7, pp. 2069-2076.

Dobrowolski, A., Romvári, R., Allen, P., Branscheid, W., & Horn, P. 2004, "Schlachtkörperwertbestimmung beim Schwein", *Fleischwirtschaft*, vol. 3, no. 2004, pp. 109-112.

Engel, B., Buist, W. G., Walstra, P., Olsen, E. V., & Daumas, G. 2003, "Accuracy of prediction of percentage lean meat and authorization of carcass measurement instruments: adverse effects of incorrect sampling of carcasses in pig classification", *Animal Science*, vol. 76, no. 2003, pp. 199-209.

Huang, J., Wium, H., Qvist, K. B., & Esbensen, K. H. 2003, "Multi-way methods in image analysis-relationships and applications", *Chemometrics and Intelligent Laboratory Systems*, vol. 66, no. 2, pp. 141-158.

Johansen, J., Egelandsdal, B., Røe, M., Kvaal, K., & Aastveit, A. H. 2006, "Comparison of different calibration models for prediction of lamb carcass composition using Computerized Tomography (CT) imaging", *Chemometrics and Intelligent Laboratory Systems*, vol. Submitted.

Jones, S. D. M., Jeremiah, L. E., Tong, A. K. W., Robertson, W. M., & Gibson, L. L. 1992, "Estimation

of Lamb Carcass Composition Using An Electronic Probe, A Visual Scoring System and Carcass Measurements", *Canadian Journal of Animal Science*, vol. 72, no. 2, pp. 237-244.

Kirton, A. H., Mercer, G. J. K., Duganzich, D. M., & Uljee, A. E. 1995, "Use of Electronic Probes for Classifying Lamb Carcasses", *Meat Science*, vol. 39, no. 2, pp. 167-176.

Nissen, P. M., Busk, H., Oksama, M., Seynaeve, M., Gispert, M., Walstra, P., Hansson, I., & Olsen, E. 2006, "The estimated accuracy of the EU reference dissection method for pig carcass classification", *Meat Science*, vol. 73, no. 1, pp. 22-28.

Romvari, R., Hancz, C., Petrasi, Z., Molnar, T., & Horn, P. 2002, "Non-invasive measurement of fillet composition of four freshwater fish species by computer tomography", *Aquaculture International*, vol. 10, no. 3, pp. 231-240.

AUTOMATIC COUINAUD LIVER AND VEINS SEGMENTATION FROM CT IMAGES

Dário A. B. Oliveira, Raul Q. Feitosa and Mauro M. Correia
Department of Electric Engineering, Catholic University of Rio de Janeiro, Rio de Janeiro, Brasil
Department of Computer Engineering, Rio de Janeiro State Universityl Brasil
Unigranrio and National Cancer Institute-INCA, Rio de Janeiro, Brasil
darioaugusto@gmail.com;raul@ele.puc-rio.br; mmauro.monteiro@gmail.com

Keywords: Medical Imaging, Liver Segmentation, Vessel Segmentation, Computed Tomography.

Abstract: This paper presents an algorithm to segment the liver structures on computed tomography (CT) images according to the Couinaud orientation. Our method firstly separates the liver from the rest of the image. Then it segments the vessels inside the liver area using a region growing technique combined with hysteresis thresholding. It separates the vessels in segments without any bifurcation, and using heuristics based on anatomy, it classifies all vessel segments as hepatic or portal vein. Finally, the method estimates the planes that best fit each of the three branches of the segmented hepatic veins and the plane that best fits the portal vein. These planes define the subdivision of the liver in the Couinaud segments. An experimental evaluation based on real CT images demonstrated that the outcome of the proposed method is generally consistent with a visual segmentation.

1 INTRODUCTION

By and large the CT data analysis is performed visually by a radiologist. This is a time consuming task, whose accuracy depends essentially on the experience of the analyst. Digital Image Processing techniques can be used to develop methods that automatically perform many of the tasks involved in the CT analysis, improving productivity and the overall accuracy.

The segmentation process is particularly arduous in abdominal CT images because different organs lie within overlaping intensity value ranges and are often near to each other anatomically. Many techniques have been proposed in the literature for the analysis of abdominal CT scans. They can be roughly divided in two main groups: model driven and data driven approaches (Masutani et al, 2005).

The blood vessel definition is an essential step in several medical imaging applications. They can be used as reference to segment different organs and structures in the human body. Kirbas et al (2004) presented a review of vessel extraction, in which many of the available techniques are described in details.

This paper presents a data driven method to segment the liver into the eight different regions proposed by Couinaud (1957), using the hepatic and portal veins position in the liver. It deals with the case of low contrast and erroneous connection between the hepatic and portal veins, as a improvement proposed in previous work (Oliveira et al., 2007).

The subsequent text is organised in the following way. Section 2 presents segmentation method in details, section 3 reports some results, and the main conclusions are presented in section 4.

2 THE 3D SEGMENTATION METHOD

The segmentation method consists of five main steps:

a) segmentation of organs and muscle tissues,
b) segmentation of the liver,
c) segmentation of the vessels within the liver ,
d) classification of segmented vessels as hepatic and portal veins, and
e) determination of Couinaud sectors.

Details of each step are presented in the next subsections.

2.1 Segmentation of Organs and Muscle Tissue

Organs and muscles tissue are the main presence in abdominal images. Typical grey values of these tissues occur around the maximum (*CM*) of the grey value histogram for the whole CT sequence.

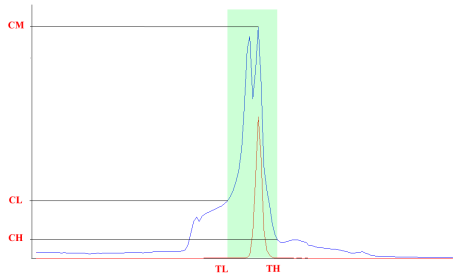


Figure 1: histogram for gray level range definition.

Figure 1 shows the histogram of a sample CT exam, the range of intensities corresponding to organs and muscles and the lower and upper limits *TL* and *TH* defining this range.

Let *CM* be the maximum CT histogram count, *TM* the corresponding intensity value, and *CL* and *CH* the counts corresponding respectively to *TL* and *TH*. It has been observed in our experiments that the ratios *RL=CL/CM* and *RH=CH/CM* do not significantly change from CT exam to CT exam. In fact these ratios lied around *RL*=0.6 and *RH*=0.2 through all our experiments.

This regularity suggests the following procedure to select the lower and higher threshold values:

a) Compute and smooth the histogram of the whole CT exam;
b) Detect the maximum histogram count *CM*;
c) Multiply *CM* by factors *RL* and *RH*, and obtain the count values *CL* and *CH*.
d) Search the smoothed histogram for the intensity values *TL* and *TH* closest to *TM* corresponding to *CL* and *CH*, such that *TL<TM* and *TH>TM*.

2.2 Liver Segmentation

The next step consists in segmenting the liver. Generally the liver appears as nearly homogeneous areas on CT slices, i.e. its intensities are restricted to a narrow grey value interval. This can be observed in Figure 1, where the histogram of pixels belonging to the liver is drawn in red over the histogram of the whole CT sequence shown in blue.

The extreme values of this interval are determined in the following way.

One image of the CT set where the liver is present is selected as the main sample and passed as an input parameter to the algorithm. Then, the largest connected component of this slice located on the upper-left side of the image (right side of the human body), is identified and its mean value on the original image is computed.

Using the pixels of organs and muscle tissue previously segmented, a new grey level range is determined following a procedure similar to the one described in subsection 2.1. The histogram count value corresponding to the liver mean value is used as the maximum count value and the range limits are calculated using as limiting ratios the value 0.8 for both cases. The threshold values obtained this way are applied to the regions selected in the previous step.

A simple procedure extracts the liver from the remaining objects. Starting on the main sample it is executed on the next adjacent slice upward and downward in the CT image set till all slices have been processed. It consists of three main steps:

a) Select the biggest object in the collection;
b) If its centroid is in the upper left quadrant of the CT image, go to step c, otherwise discard this object from the collection and go back to step a;
c) If the selected object is connected to another object of an adjacent slice previously classified as liver, classify it as liver, otherwise discard the object from the collection and go back to step a;

Clearly the first iteration does not pass through step c and the object selected in step b is set as liver directly.

2.3 Vessel Segmentation

Having segmented the liver, and considering only the region delimited by this organ, we select a threshold *VH*, such that the intensities above it identify unambiguously the vessels inside the liver. A second threshold *VL* (*VL<VH*) is further selected such that intensities below it clearly indicate liver parenchyma.

These two threshold values define three ranges of pixel intensities, namely:

- the strong vessel range, defined by intensities above *VH*,
- the weak vessel range, comprising intensities between VL and VH, and

- the liver tissue range, covering intensities below VL.

The construction of the vessel tree is performed by a region growing approach consisting of the following basic steps:

a) Build the weak vessel object set defined by the pixels with values above VL.
b) Build the strong vessel object set defined by the pixels with values above VH.
c) Take the strong vessel set computed in the preceding step as the initial vessel tree estimate, and add to it all objects of the weak vessel set connected to it.
d) Repeat the previous step using the current vessel tree estimated until it stops growing.

We searched appropriate values for *VL* and *VH* manually through many experiments using different CT sequences. We observed that the histogram counts for the manually selected values stayed at a roughly constant ratio to the intensity corresponding to the maximum count.

Considering *NM* the maximum liver histogram count, and *NL* and *NH* the counts corresponding respectively to *VL* and *VH*, the ratios *rl=NL/NM* and *rh=NH/NM* do not significantly change from CT exam to CT exam. These ratios were determined experimentally as *rl*=0.5 and *rh*=0.2.

Based on this regularity the following procedure is proposed to select the lower and higher threshold values:

a) Compute and smooth the histogram of the image region inside the liver;
b) Detect the maximum histogram count NM and the corresponding intensity VM.
c) Multiply NM by the ratios rl and rh, and obtain the count values NL and NH.
d) Search the smoothed histogram for the intensity values VL and VH corresponding to NL and NH, whereby both VL and VH are greater than VM.

2.4 Classification of Portal and Hepatic Veins

The hepatic and portal veins appear as separate three dimensional objects in most CT exams. However, sometimes these veins touch to each other on some CT slice, what may lead to identifying them as a single object. In such case the Couinaud segmentation becomes not possible.

This subsection describes a method to correctly segment the veins even when they touch in some CT slice.

Firstly, the method separates the vessel objects segmented previously in connected components, hereafter called objects, performing the following steps:

a) The first slice S1 containing any object is labelled.
b) The area projected by each object in S1 on the next adjacent slice S2 is verified. If it intersects only one object, the same label is set to the object in S2. If it intersects more than one object, new labels are created for each intersected object in S2.
c) Step b is repeated until all objects in the CT sequence are labelled.

As result vessels segments are obtained whose extremes are determined by bifurcations, as shown in Figure 2-b.

A second procedure is performed to classify these vessel segments as hepatic or portal vein. Based on knowledge of the anatomy, the following simple algorithm is proposed. It consists of six steps:

a) The first object identified on the top slice is selected.
b) If it is divided in three other objects in the next adjacent slice, it is classified as hepatic vessel, otherwise it is discarded and other object is selected on the top slice until this condition is reached.
c) For each of the three objects identified as hepatic branches, the next adjacent slice is analysed. The object with the largest intersection area is selected as continuation of the respective hepatic branch.
d) Step c is repeated recursively for each hepatic branch until no other segment can be merged to the hepatic vessel tree. At the end of this step, the major hepatic vessels have been identified.
e) The vessel segments not assigned to the hepatic vessel tree up to step d are examined and the largest 3D connected component is labelled as the portal vein.
f) Non classified segments which are connected to the hepatic vessel tree are merged to it.

In figure 2-a the hepatic and portal veins are shown as a single object because they touch on some CT slice. Figure 2-b shows in different colours several independent segments delimited by each bifurcation identified during the classification process. Figure 2-c shows the final result, where the hepatic and portal veins appear as separated vessel trees.

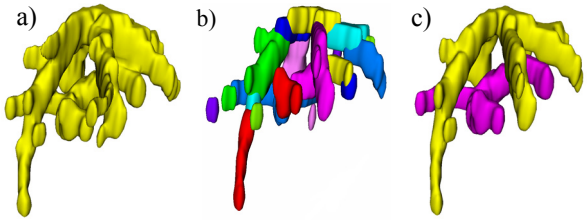


Figure 2: Vessel segmentation results – (a) the portal and hepatic veins (b) the independent vessel segments determined by bifurcations (c) the portal and hepatic veins as separated vessel trees.

2.5 Segmentation of Couinaud regions

The Couinaud paradigm divides the liver into eight independent segments each one having its own vascular inflow, outflow, and biliary drainage. Because of this division into self-contained units, each can be removed without damaging those remaining.

Our method estimates the subdivision of the liver in the eight Couinaud segments, by fitting planes to the portal vein, and to each of the hepatic vein branches. To separate the three main branches of the hepatic vein we apply the k-means algorithm on the 3 dimensional coordinates of the pixels identified in the preceding step as belonging to the hepatic vein. It is assumed that there are three clusters. A restriction for singleton value is imposed so as to guarantee that no cluster will be empty. This leads to three different objects corresponding to each branch of the hepatic vein.

Then, a least squares based procedure determines the four planes that best fit the points of each branch of the hepatic vein and the portal vein segmented before. These planes divide the liver in the Couinaud segments.

3 RESULTS

A software prototype implementing the proposed method has been built for validation purpose. It also implements both the surface and volumetric visualization of the internal liver structures. It receives as input the segmented structures of each image slice and the thickness of the CT slices available in the DICOM image file header.

Figure 3 shows an example of segmentation result produced by the proposed procedure as a 3D surface which can be visualized within our prototype. It is possible to observe the hepatic vein and the portal vein respectively in blue and red, and the Couinaud segments in different colours are also present. It can be observed that the Couinaud segments are divided according to the veins orientation.

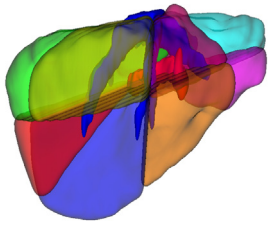


Figure 3: 3D models of segmented liver structures.

Experiments performed on seven different CT sequences have shown that the results produced by the proposed method are consistent with the visual perception of a specialist.

4 CONCLUSIONS

This work proposes an algorithm to segment the liver in computer tomography (CT) images according to the Couinaud classification.

Experiments conducted on a software prototype of the proposed algorithm upon 7 CT produced results consistent with the visual perception. The method has the potential of becoming a useful tool in various applications. It can be used to generate 3D liver representations to aid visual diagnostic and surgery planning. Shape attributes other than volume may also be measured from the 3D model and explored in Computer Aided Diagnostic environments.

The assessment of segmentation accuracy is a major concern in the continuation of this work.

REFERENCES

Couinaud, C. “Le Foie: Etudes Anatomiques et Chirurgicales. Masson”. Paris. 1957.

Kirbas, C. and Quek, F. “A review of vessel extraction techniques and algorithms.” ACM Comput. Surv. 36, 2 (Jun. 2004), 81-121. 2004.

Masutani Y., Uozumi K., Akahane M. and Ohtomo K., “Liver CT image processing: A short introduction of the technical elements”, E. Journal of Radiology, Vol 58, Liver Lesions, Pages 246-251. 2006.

Oliveira, D.A.B., Mota, G.L.A., Feitosa, R.Q. and Nunes, R.A., "A region growing approach to pulmonary vessel tree segmentation using adaptive threshold". CompIMAGE proceedings, pp. 319-324, 2006, Porto.

BIOSIGNALS ANALYSIS AND ITS APPLICATION IN A PERFORMANCE SETTING

Towards the Development of an Emotional-Imaging Generator

Mitchel Benovoy, Jeremy R. Cooperstock
Centre for Intelligent Machines, McGill University, 3480 University St., Montreal, Canada
benovoym@cim.mcgill.ca, jer@cim.mcgill.ca

Jordan Deitcher
Director, E-Motion Project
liminal@total.net

Keywords: Biosignals, Pattern Recognition, Signal Processing, Emotions, Emotional Imaging, Instrument, Performance Art.

Abstract: The study of automatic emotional awareness of human subjects by computerized systems is a promising avenue of research in human-computer interaction with profound implications in media arts and theatrical performance. A novel emotion elicitation paradigm focused on self-generated stimuli is applied here for a heightened degree of confidence in collected physiological data. This is coupled with biosignal acquisition (electrocardiogram, blood volume pulse, galvanic skin response, respiration, phalange temperature) for determination of emotional state using signal processing and pattern recognition techniques involving sequential feature selection, Fisher dimensionality reduction and linear discriminant analysis. Discrete emotions significant to Russell's arousal/valence circumplex are classified with an average recognition rate of 90%.

1 INTRODUCTION

Emotion classification based on *external* data collection schemes, such as speech analysis and facial-expression recognition from images has been studied extensively. The literature offers numerous examples of relatively acceptable recognition rates (Black et al., 1995; Lyons et al., 1999; Bartlett et al., 1999; Ververidis et al., 2004). However, because these systems require sensors, such as cameras or microphones, focused directly on the subject, they are restrictive in terms of movement and problematic in terms of signal interference from other devices. Moreover, video analysis methods tend to encourage exaggerated physical expressions of emotion that are often artificial and uncorrelated with the actual emotion being experienced by the individual.

In contrast, biosignal analysis, based on skin surface sensors worn by the user, may be a more robust and accurate means of determining emotion. This is because the signals correspond to internal physiology, largely related to the autonomous nervous and limbic systems, rather than to external expressions that can be manipulated easily. However, emotional state recognition by means of biosignals analysis is also problematic. This is due in part to the movement sensitivity of physiological sensors to such signals as electrocardiograms (ECG) and galvanic skin response (GSR). Muscle contractions are induced by electrical neural impulses, which in turn are picked up by the devices designed to measure differences in electrical potential. These may cause noise in the form of signal fluctuations. Furthermore, despite the evidence from psychophysiology suggesting a strong correlation between human emotional states and physiological responses (Watanuki et al., 2005; Cacioppo et al., 1990), determining an appropriate mapping between the two is nevertheless non-trivial.

Our interest in these techniques differs significantly from previous work. Rather than recording and classifying how people respond to external stimuli such as culturally meaningful images, sounds, film clips, and text, we are in the process of developing a biometrically driven multimedia instrument, one that enables a performer

to express herself with artistry and emotional cohesiveness. The goal is to provide a rich, external manifestation of one's internal, otherwise invisible, emotional state. With training, it is our hope that the resulting system, one that is coupled to the performer's emotional intentionality rather than to external gestures, can become as expressive and responsive as a fine musical instrument. Thus, rather than attempt to recognize and label human emotional states, our goal is to investigate the mapping of these states to expressive control over virtual environments and multimedia instruments.

From an artistic perspective, the instrument interface should support the articulation of emotion in a meaningful manner, with acuity and subtlety, allowing it be played with sensitivity and nuance.

We see the development of this instrument as a two-stage process. The first phase, described in this paper, deals with the question of emotion capture, that is, extracting meaningful data from the range of sensors available to us.

The second stage, which we discuss briefly in Section 5, relates these signals to the output of the instrument and how it is designed to be used in a performance setting. Because the instrument is ultimately a highly enriched biofeedback device, a performer's response to anything and anyone she encounters, including the audience, instantly manifests all around her. To bring it under her control, she must first *compose* herself. This involves using the instrument as a feedback device to return to a neutral state from which all emotions are potentially accessible. Once she has done so, she can put the instrument to its true use, directing her emotions outward in the act of creative composition.

The remainder of this paper is organized as follows. Our emotion elicitation method, used to gather the physiological data, is described in Section 3. Next, the recognition engine, including feature selection, reduction and classification, is described in Section 4. Finally, Section 5 concludes with a discussion of some future avenues for research.

2 RELATED WORK

Ekman's emotion classification scheme (Ekman, 2005) included six principal, discrete and universal classes of affect: anger, joy, fear, surprise, disgust and sadness. Russell's arousal/valence circumplex (Posner et al., 2005) introduced a continuous, analog mapping of emotions based on a weighted combination of arousal intensity and emotional valence (negative to positive). Figure 1 depicts this two-dimensional space with an example set of emotions.

For our purposes, both types of representations are useful for "playing" the *instrument* represented by the high-level schematic of Figure 2: discrete states serving as coarse control, with the analog input driving fine-tuned and subtle variations.

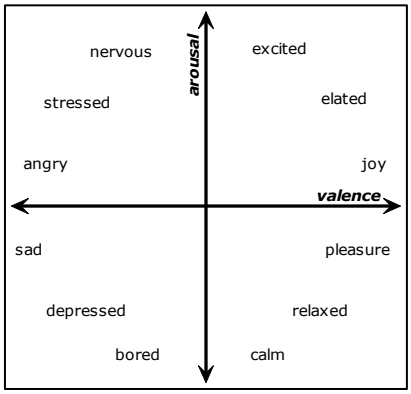


Figure 1: Russell's arousal/valence circumplex (reproduced from Posner et al., 2005).

Previous studies have demonstrated that emotional arousal and valence stimulate different brain regions (Anders et al., 2004) and in turn affect peripheral systems of the body. Significant physiological responses to emotions have been studied, showing, for example, measurable changes in heart rate and phalange temperature in fearful, angry and joyful states (Ekman et al., 1983).

Emotional state recognition using physiological sensors has been investigated by others. Picard (Picard et al, 2001) obtained good recognition results (81.25% accuracy) on eight emotions using one subject stimulated with personally selected images and four physiological sensors: blood volume pulse (BVP), galvanic skin response (GSR), electromyograph, and respiration). Our results, restricted to four emotions, are similar, but the critical difference between our approaches is the elicitation process. While Picard uses images to elicit emotion, we focus on an involved self-generation of affective states. This, we believe, has important implications for real-world theatrical performance, where emotions are continuously varying as opposed to discrete. Capturing the subtle dynamics of emotion is vital to attaining the cognitive and emotive skills required for mastering control of the instrument.

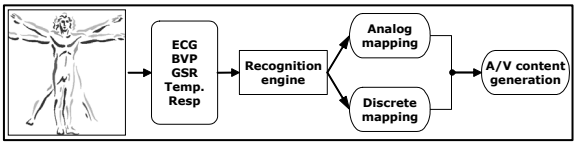


Figure 2: Biosignals-driven Emotional-Imaging Generator.

3 EMOTION ELICITATION

As noted above, we are primarily interested in how self-generated emotional states can be mapped through biosignal analysis to the proposed instrument. Clearly, the performer must be skilled in the art of accessing and articulating emotion. Just as with learning any musical instrument, feedback must be provided that connects her meaningfully both with the appropriate skill level and emotional experience.

As a first step in investigating these issues, we want to capture biosignal data of maximum possible validity. Gaining access to the ground truth of human emotion remains an elusive goal. Nevertheless, we can obtain a better labelled set of input than that available through generic stimuli, as used by other researchers. To do so, we interact directly with the experimental subject to *generate* the stimuli. This avoids the potential problems, articulated by colleagues, of subjects not responding to a particular stimulus as expected, or verbally expressing an emotion "they think the stimulus is supposed to evoke."

Of course, this necessitates that the stimulus be highly personalized and subjective. The benefit is the potentially greater physiological validity of the recorded data that is then used for training (or calibrating) our system. As seen in the results of Section 4, we succeed in obtaining an encouraging correct classification result over four emotions of 90%.

3.1 Experimental Subject

To maximize the validity of our experimental data, we worked with a professional method actor, who was guided by one of the authors (Deitcher), an experienced theatre director. Our subject has had the opportunity to methodically investigate an extraordinarily wide array of characters and situations. Effective emotional solicitation from someone with this kind of experience and flexibility requires the sensitivity to anticipate relevant emotional connections. It also requires the ability to ask the questions and define the exercises that will allow these emotions to emerge. In the broadest of terms, by having the actor play scenes, sing songs, follow guided visualizations and remember events from her own life, we were able to elicit a large and complex range of emotional landscapes. Her focused intentionality was responsible for engendering a high degree of confidence in the collected physiological data.

3.2 Experimental Data Collection

Experiments were conducted in a quiet, comfortable lab environment. The subject either remained seated or standing and was instructed to limit her body movement to minimize motion artefacts in the collected signals. The biosignals were recorded using Thought Technology's *ProComp Infiniti* biofeedback system using five sensor channels: GSR, ECG, BVP, phalange temperature and respiration, all sampled at 256 Hz. Each trial was also videotaped with a synchronization signal to align the video recording with the biosignals.

3.3 Data Types

Two types of data were recorded: discrete emotional states and the responses to complex emotional scenarios. Typical trial times of 60 and 300 seconds were used for each type of data, respectively. A fifteen-minute break was taken between each trial so that the subject could return to her *baseline*, emotionally relaxed state.

The discrete class of data afforded a simple labelling of emotions, as expressed by the subject during each trial. These were used primarily for classifier training and validation. During these experiments, the subject was asked to experience four emotional states in turn (joy, anger, sadness, pleasure), while vocalizing what she was feeling. A post-trial questionnaire was used to determine a subjective assessment of the intensity of the sensed emotion, on a numeric scale from one to five. Twenty-five trials of each of the four emotions were recorded.

For the complex scenarios, data segments were recorded while the subject acted out "scenes" of fluid and varying emotional states. Such experiments will be used to study the body's psychophysiological responses during emotional transitions. These scenarios are theatrically dynamic, and thus meaningful in investigating the performance possibilities of our proposed instrument.

4 RECOGNITION ENGINE

Our preliminary investigations deal only with the classification of discrete emotional states to validate our paradigm of emotion elicitation, described in the previous section. The recognition engine comprises two main stages: biosignals processing and classification, both implemented in Matlab.

The emotional state recognition system utilizes five physiological signals: electrocardiogram (ECG), GSR, BVP, respiration and phalange temperature. We employ digital signal processing and pattern recognition, inspired by statistical techniques used by Picard. In particular, our use of *sequential forward selection* (a variant of sequential floating forward selection), as used by Picard, choosing only classifier-optimal features, followed by *Fisher dimensionality reduction*, are similar. For the classification engine, however, we implemented linear discriminant analysis rather than the maximum *a posteriori* used by Picard.

4.1 Biosignal Processing

The raw, discrete biosignals go through four steps to produce classifier-ready data, as shown in Figure 3.

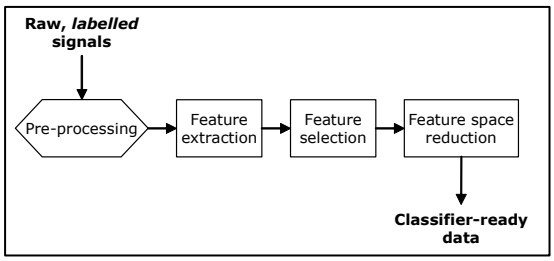


Figure 3: Biosignal processing engine.

4.1.1 Pre-Processing

Emotionally relevant segments of the recordings that are free of motion artefacts are hand-selected and labelled with the help of the video recordings and responses to the questionnaire. High-frequency components of the signals are considered to be noise and filtered with a Hanning window (Oppenheim, 1989).

4.1.2 Feature Extraction

We extract six common statistical features from each type of the noise-filtered biosignals, of size N ($X_n, n \in [1 \dots N]$), and its first and second derivatives:

1. Filtered signal mean:

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X_n \quad (1)$$

2. Filtered signal standard deviation:

$$\sigma_x = \left(\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu_X)\right)^{1/2} \quad (2)$$

3. Filtered signal mean of absolute value of the first difference:

$$\delta_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|X_{n+1} - X_n\right| \quad (3)$$

4. Normalised signal mean of absolute value of the first difference:

$$\tilde{\delta}_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{X}_{n+1} - \tilde{X}_n\right| = \frac{\delta_x}{\sigma_x} \quad (4)$$

5. Filtered signal mean of absolute value of the second difference:

$$\gamma_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|X_{n+1} - X_n\right| \quad (5)$$

6. Normalised signal mean of absolute value of the second difference:

$$\tilde{\gamma}_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{X}_{n+1} - \tilde{X}_n\right| = \frac{\gamma_x}{\sigma_x} \quad (6)$$

Where $\tilde{X}_n$ represents the normalised signal (zero-mean, unit variance):

$$\tilde{X}_n = \frac{X_n - \mu_x}{\sigma_x} \quad (7)$$

In addition to the previous features, used for each biosignal, other signal-specific characteristics are computed. These include, for example, heart rate mean, acceleration/deceleration and respiration power spectrum at different frequency bands. Combining the statistical and signal-specific characteristics, a total of 225 features are thus computed from the five types of biosignals.

4.1.3 Automatic Feature Felection

Feature selection is a method widely used in machine learning to select a subset of relevant features in order to build robust learning models.

The aim is to remove most of the redundant and irrelevant features from the data to alleviate the often detrimental effect of high dimensionality and to improve generalization and interpretability of the model.

The greedy sequential forward selection (SFS) algorithm is used to form automatically a subset of the best n features from the original large set of m ($n < m$). SFS starts with an empty feature subset and on each iteration, exactly one feature is added. To determine which feature to insert, the algorithm tentatively adds to the candidate feature subset one that is not already selected and tests the accuracy of a k-NN classifier built on this provisional subset. A feature that results in the highest classification accuracy is permanently included in the subset. The process stops after an iteration where no feature addition causes an improvement in accuracy. The resulting feature set is now considered optimal.

The k-NN classifier used here classifies a novel object r by a majority of "votes" of its neighbours, assigning to r the most common class among its k nearest neighbours, using the Euclidean distance as metric. This type of classifier is chosen because it is a simple and efficient performance criterion for feature selection schemes and is considered more robust than using a single measure of distance, as is the case for many feature selection schemes. It was found through iterative experimentation using $k \in [1,9]$, that a value of $k = 5$ resulted in the best possible selected feature subset.

4.1.4 Feature Space Reduction

Fisher dimensionality reduction (FDR) seeks an embedding transformation such that the between-class scatter is maximized and the within-class scatter is minimized, resulting in a low-dimension representation of optimally clustered class features. FDR is shown to produce optimal clusters using $c - 1$ dimensions, where c is the number of classes. However, if the amount of training data or the quality of the selected feature subset is questionable, as is the case in many machine learning applications, the theoretically optimal dimension criterion may lead to an irrelevant projection which minimizes error in the training data, but performs badly with testing data (Picard et al., 2001). In our case, a two-dimensional projection resulted in an overall best classification rate using linear discriminant analysis (LDA) to sequentially test with dimensions $d \in [1,3]$. Figure 4 demonstrates the class clustering of four emotional states: joy, anger, sadness, pleasure (JO, AN, SA, PL), projected on a 2D Fisher space during one of the validation steps. The four emotions were chosen given that they lie in different quadrants of Russell's arousal/valence circumplex (Figure 1).

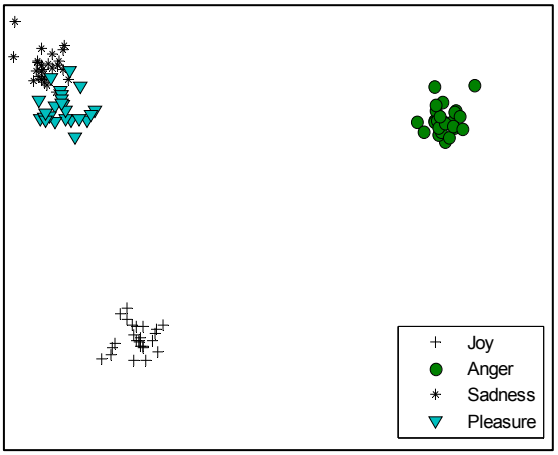


Figure 4: 2D Fisher projection (4 classes).

4.2 Biosignal Classification

Three popular classification schemes were tested to classify the four emotional states: LDA, k-NN ($k \in [1,9]$) and multilayer perceptron (MLP). LDA was found to outperform both the best k-NN (k = 7) and MLP by 4% and 11%, respectively. LDA builds a statistical model for each class and then catalogues novel data to the model that best fits. We are thus concerned with finding which discriminant function best separates the emotion classes. LDA finds a linear transformation Φ of the x and y axes (8) that yields a new set of values providing an accurate discrimination between the classes. The transformation thus seeks to rotate the axes with parameter v so that when the data is projected on the new axes, the difference between classes is maximized.

$$\phi = v_1\mathbf{x} + v_2\mathbf{y} \qquad (8)$$

Due to the small feature dataset size, leave-one-out cross-validation was used to test the classification scheme. This involves using a single item of the set as the validation data, and the remaining ones as training data. This process is repeated until each item in the dataset is used once as the validation data. At each iteration, SFS and FDR are applied to the new training set and the parameters found (selected features and Fisher projection matrix) are applied to the test set. The mean classification rate is computed using the result

produced at each step. Using this method, our biosignal classification system produced an average recognition rate of 90% on the four emotional states. Table 1 shows the confusion matrix for the classification.

Table 1: LDA classifier confusion matrix.

I/O	JO	AN	SA	PL	%
JO	**0.96**	0	0	0.04	**96**
AN	0	**1.00**	0	0	**100**
SA	0.04	0	**.92**	0.04	**92**
PL	0.12	0	0.16	**0.72**	**72**

5 CONCLUSIONS

A novel emotion elicitation scheme based on self-generated emotions is presented, engendering a high degree of confidence in collected, emotionally relevant, biosignals. Discrete state recognition via physiological signal analysis, using pattern recognition and signal processing, is shown to be highly accurate. A correct average recognition rate of 90% is achieved using sequential forward selection and Fisher dimensionality reduction, coupled with a Linear Discriminant Analysis classifier.

We believe that the high classification rate is due in part to our use of a professional method actor as test subject. It is speculated that normal subjects would lead to lower rates because of the high variability of emotion expressivity across a large population pool. It is an avenue of research for us to test the generalization of this type of machine-based emotion recognition.

Our ongoing research also intends to support *real-time* classification of discrete emotional states. Specifically, continuous arousal/valence mappings from biosignals will drive our emotional-imaging generator for multimedia content synthesis and control in a theatrical performance context. In addition, we are exploring the therapeutic and performance training possibilities of our system. Because what we are building is fundamentally an enriched biofeedback device, we anticipate applications ranging from stress reduction for the general population to the generation of concrete emotional expression for those with autism or other communication disorders.

ACKNOWLEDGEMENTS

The authors wish to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) New Media Initiative and the Centre for Interdisciplinary Research in Music Media and Technology at McGill University for their funding support for this research. Special thanks are also due to Laurence Dauphinais, who gave many hours of her time and her artistic insight, and to Thought Technology Ltd., which provided the acquisition hardware and software used in this research.

REFERENCES

Anders S., Lotze M., Erb M., Grodd W., Birbaumer N., 2004. *Brain activity underlying Emotional valence and arousal: A response-related fMRI study*. Human Brain Mapping, Vol. 23, p. 200-209.

Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J., 1999. *Measuring facial expressions by computer image analysis.* Psychophysiology, Vol. 36, p. 253-263.

Black, M.J., Yacoob, Y., 1995. *Recognizing facial expressions in image sequences using local parameterized models of image motion*. ICCV.

Cacioppo, J., Tassinary, L.G., 1990. *Inferring psychological significance from physiological signals.* American Psychologist, Vol 45, p. 16-28.

Ekman, P., Levenson, R.W., Friesen, W.V., 1983. *Autonomic Nervous System Activity Distinguishes Between Emotions.* Science, 221 (4616), p. 1208-1210.

Ekman P., 2005. *Emotion in the human face*, Cambridge University Press, p. 39-55.

Lyons, M. Budynek, J., Akamatsu, S. 1999. *Automatic Classification of Single Facial Images.* IEEE PAMI, vol. 21, no. 12.

Oppenheim, A.V., Schafer, R.W., 1989. Discrete-Time Signal Processing, Englewood Cliffs, N.J.: Prentice-Hall.

Picard, R.W., Vyzas, E., Healey, J., 2001. *Toward machine emotional intelligence: analysis of affective physiological state*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 23, Issue 10, p. 1175 – 1191.

Posner J., Russell J.A., Peterson B.S., 2005. *The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology*. Development and Psychopatholy, p. 715-734.

Ververidis, D., Kotropoulos, C., Pitas, I., 2004. *Automatic emotional speech classification*, IEEE ICASSP.

Watanuki S., Kim Y.K., 2005. *Physiological responses induced by pleasant stimuli*. Journal of Physiological Anthropology and Applied Human Science, p. 135-138.

ANALYSIS ALGORITHMS FOR A FIRST-AID SENSOR
Detecting Vitality Parameters such as Pulse and Respiration

Daniel Wettach, Marc Jaeger, Armin Bolz
Institute of Biomedical Engineering, University of Karlsruhe, Hertzstr. 16, Karlsruhe, Germany
Daniel.Wettach@ibt.uni-karlsruhe.de, Marc.Jaeger@ibt.uni-karlsruhe.de, Armin.Bolz@ibt.uni-karlsruhe.de

Timur Oezkan
Department of Anaesthesiology, University Hospital Heidelberg, Germany
timur.oezkan@gmx.de

Keywords: Physiological processes, First-aid sensor, real-time sensor system, wearable sensor and system, computational intelligence, emergency situation.

Abstract: In this paper the software algorithms necessary to analyze the signal provided by a first-aid sensor system that detects pulse and respiration at a single point are described. In an opinion poll four of five inexperienced first responders were interested in using this kind of system as support in emergency situations. Especially the intelligent detection of respiration is hardly popular today and in most cases only possible offline. The software also controls several visual indicators that assist the first aider in quickly determining the state of the patient.

1 INTRODUCTION

In emergency situations like an accident a first aider has to decide immediately if resuscitation of the victims needs to be initiated. The know-how of many first aiders is not sufficient or their uncertainty too great to be able to make this often vitally important decision quickly (Sefrin, 2006). Our aim is to develop a cheap and portable sensor system that is easy and fast to use to support the first-aider with their vital decision. It detects if pulse and respiration of the victim work normally and provides the first aider with a visual aid. The sensor has to be placed on the neck of the victim and after a few seconds it starts to display information about pulse and respiration.

2 SIGNAL PROCESSING

2.1 Medical Concept

It is our aim to develop a sensor which measures not only the pulse but also the frequency of respiration, that is why the detection of electrical body signals e.g. like an ECG is not sufficient. One solution is to measure mechanical changes of the tissue surface which are caused by both pulse beat and respiration. As possible measuring point for the sensor a point near the clavicle is estimated to deliver the best results because that point is close to both the subclaviar artery and the trachea. This point is shown in Fig. 1.

Figure 1: Possible measuring point for the sensor.

2.2 Signal Acquisition

The measuring of changes in the tissue surface is mainly done by detecting changes in distance between the sensor and the tissue surface. Normally the measuring of distances can be achieved by using an LC oscillator. Due to capacitive coupling the resonance frequency of the oscillator changes if the

distance to any organic object in- or decreases. This change in resonance frequency is reflected in a change of voltage. This principle is used in homes in the form of touchless switches. However, since the changes in the tissue surface caused by pulse beat and respiration are very small, the use of classical LC oscillators is - apart from a few exceptions - not possible. Therefore a new nonlinear oscillator has been developed. With its help it is possible to measure changes in submillimeter range. Because of this high sensitivity even little muscle contractions will be detected by the sensor. As some of these minor contractions do belong to neither pulse nor respiration they have to be eliminated. It is necessary to filter and process the signal to eliminate these artifacts. An example for the measured signal is shown in Fig. 2. For more technical information please refer to (Jaeger, 2007).

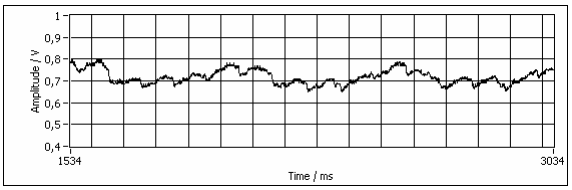


Figure 2: Signal measured by the sensor.

2.3 Algorithms

First the signal is band-filtered with hard coded cut-off frequencies covering the whole spectrum of possible pulse and respiration frequencies. In the next step a low resolution FFT of this filtered signal is performed. In extreme cases the respiration frequency can be higher than the pulse frequency. Due to this there are four possible frequency bands (Fig.3):

- the whole frequency band from the lowest possible frequency of the respiration to the highest possible frequency of the pulse (WF)
- the lower frequency band of respiration from the lowest possible frequency of the respiration to the lowest possible frequency of the pulse (LF),
- the upper frequency band of pulse, from the highest possible frequency of the respiration to the highest possible frequency of the pulse (HF)
- and the overlapping band from the lowest possible frequency of the pulse to the highest possible frequency of the respiration (MF).

By calculating relations between the different amplitudes, the frequency parts with high intensity can be discerned. Normally there will be only two frequency parts with high intensity representing pulse and respiration.

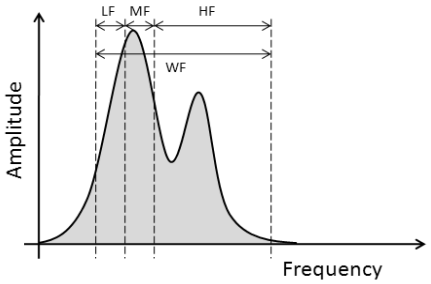


Figure 3: Spectrum of the filtered signal with four frequency bands.

If there are more than two such frequency parts, there has to be a measurement error. If no error is detected the frequency parts defined above can be analyzed further. If there is a frequency part of high intensity in the lower frequency band of respiration, there may be only one such frequency part in the overlapping band - the one of the pulse. If such a frequency part is additionally detected in the upper frequency band of the pulse there cannot be any in the overlapping band. As soon as these conditions are met, the frequencies for pulse and respiration are approximately determined. With the detected frequencies it is possible to define narrow cut-off frequencies for pulse and respiration. With these cut-off frequencies the input signal is filtered again, once for pulse and once for respiration.

After this step there are two rather clear signals which are relatively free of errors. With these filtered signals it is possible to determine more exact frequencies for pulse and respiration. Therefore the filtering is designed to adapt which means that the cut-off frequencies of the fine filters are always readjusted using the frequencies determined by the fine filtered signals. Additionally the approximated frequencies of pulse and respiration gained from the roughly filtered input signal are used to validate the adjustment of the cut-off frequencies for the fine filters. The determination of the frequency of the fine filtered signal works as follows: On the one hand zero-crossings are counted using a hysteresis, on the other hand a high resolution FFT for the filtered signals is processed. Both results undergo a plausibility check.

The frequency determination by counting zero-crossings is done by using hysteresis. A zero-crossing is only interpreted as such if the signal does not only drop below a lower hysteresis border but

then also rises over an upper hysteresis border. The point in time in which the signal rises over the upper hysteresis border is used as zero-crossing. The time difference between two such points can be interpreted as the periodic time of the signal. The reciprocal of this periodic time is the frequency of the signal. To get a more stable frequency value the last two periodic times determined in this way are averaged. If the signal does not pass a hysteresis border for a certain period of time its quality is assumed to be insufficient and the frequency value will be set to zero.

When the frequency is determined by using a high resolution FFT the frequency part with the highest amplitude is used as result. To get a more reliable value the algorithm averages out the last five determined frequency values.

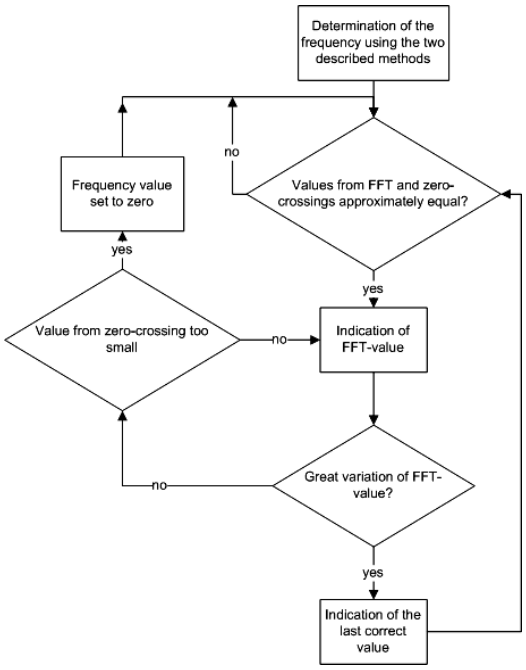


Figure 4: Logical organization of the plausibility check.

To eliminate remaining errors a plausibility check is done using the determined frequency values by counting zero-crossings and FFT. Fig. 4 shows the logical organization of the plausibility check.

As soon as both methods - counting zero-crossings and FFT - calculate approximately the same frequency value the signal is assumed to be free of errors and the value calculated by FFT will be indicated as final result. If this value deviates too much from the previous shown result an error will be assumed. In this case the last correct value prior to the error is indicated until both methods once again calculate an approximately equal value. If the value determined by counting zero-crossings drops below a predetermined border, the signal is assumed to be too weak and the final frequency value will be set to zero.

Fig. 5 shows the chronological sequence of possible results of the two methods counting zero-crossings and FFT and the corresponding result of the plausibility check (frequency line in Fig. 5).

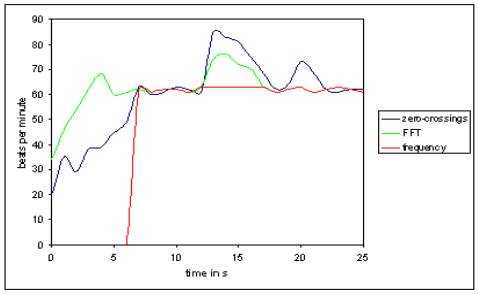


Figure 5: Possible chronological result of the plausibility check.

2.4 Visualization

The sensor uses LEDs to indicate the state of a patient. For both pulse and respiration there is a LED showing if the respective function works normally. If either pulse or respiration are abnormal, it will be shown by an additional LED. A fourth LED indicates that the sensor system is working correctly to avoid any delay due to unrecognized malfunctions. The sensor with its LEDs is shown in Fig. 6.

Figure 6: The Sensor with the four LEDs.

If the LEDs indicate that pulse and/or respiration does not work normally and the patient does not show any reactions to external stimulation the first aider has to initiate cardiopulmonary resuscitation (CPR).

3 APPLICATION

The following diagram (Fig. 7) shows the sequence from the arrival of a first aider at an accident up to the corresponding action he has to take.

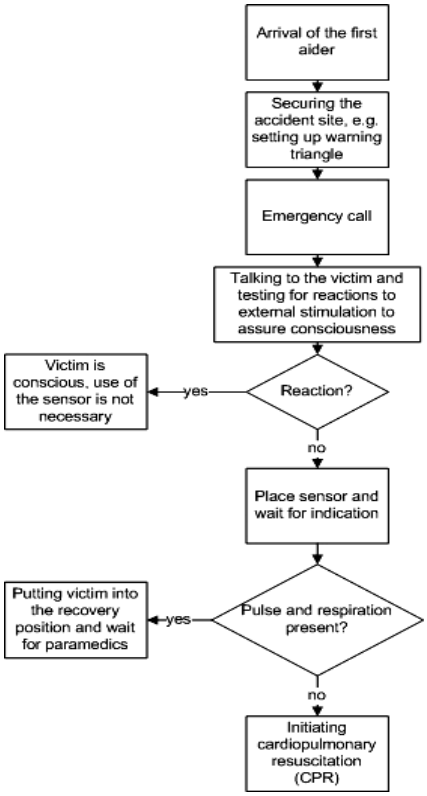


Figure 7: First aider mode of operation.

4 RESULTS

The developed algorithms have been tested with several signals recorded by the sensor. The time it took the software to detect each the pulse and the respiration signals has been put in relation to the overall time of the measurement. The results are shown in Table 1. The sensor is able to detect pulse in 91.15% and respiration in 81.15% of the time the sensor is active. The timeframe during which pulse and respiration are not detected is mostly at the beginning of the measurement because at least one or two cycles are necessary to calculate useful values. The lower percentage for respiration detection can be explained with the lower breathing frequency due to which the initial cycles take longer to complete.

Table 1: Test cases for pulse and respiration.

	Signal length in s	Pulse detected in %	Respiration detected in %
75.txt	132,92	92,33	87,71
84.txt	191,55	97,49	92,69
87.txt	97,90	95,86	85,41
90.txt	80,63	94,51	78,69
137.txt	76,13	96,69	79,64
190.txt	119,59	97,12	71,99
carotis.txt	29,22	83,06	57,56
kieferwinkel.txt	169,61	97,70	91,16
sternoclaido.txt	199,36	83,74	82,68
subclavia.txt	74,99	73,00	84,00

5 DISCUSSION

The results shown in Table 1 are not satisfying because they were not verified by comparing them to actual pulse and respiration data detected by other means. In addition the sensor has to be tested in extreme situations for instance on board of rescue helicopters and ambulances again while comparing the sensor data to actual pulse and respiration data. According to these test results - like possible deviations of the sensor detected data from the actual pulse and respiration - the algorithms will have to be improved. The main task in the future will be to find ways to correct inaccuracies caused by small movements. This could be done for example by integrating a neural fuzzy system into the software that can use additional criteria to ascertain that the calculated values are correct.

Another possibility is to include an acceleration sensor in the system that is able to detect certain movements and to correct the data accordingly.

REFERENCES

P. Sefrin, 2006, *Bei der Reanimation gilt ab sofort: Zuerst 30 mal das Herz massieren und dann zweimal beatmen*, Aerztezeitung

M. Jaeger, M. Mueller, D. Wettach, T. Oezkan, J. Motsch, T. Schauer, R. Jaeger, A. Bolz, 2007, *First-Aid sensor system: New methods for single-point detection and analyzation of vitality parameters such as pulse and respiration,* IEEE EMB, ISBN 1-4244-0788-5

A FULLY AUTOMATIC RED-EYES DETECTION AND CORRECTION ALGORITHM BASED ON UNIFORM COLOR METRIC AND BINOCULAR GEOMETRIC CONSTRAINT

Chun-Hsien Chou[*], Kuo-Cheng Liu[*,**] and Shao-Wei Su[*]
[*] *Department of Electrical Engineering, Tatung University, Taiwan*
[**] *Foreign Language and Information Educating Center, Taiwan Hospitality and Tourism College*

Keywords: Red-eye, Digital photography, Digital cameras, Uniform color space, Color difference.

Abstract: Red-eye is a highly objectionable defect that often occurs in digital images taken with a flash by modern small cameras. Although many red-eye reduction algorithms were proposed and equipped in most of the digital cameras, none of these algorithms is effective enough. In this paper, an algorithm for automatic detection and correction of red-eyes is proposed. The color detector based on uniform color metric is developed to locate regions of major colors including red-eye color and skin tone. The structure of major colors is adopted to locate candidate red-eye regions. The geometric relationship between the dimension of the human pupil and binocular distance is employed to eliminate most false positives (image regions that look like red-eyes but are not). More than one pairs of red-eyes snapped in different view angles are detected by the proposed algorithm. Detected red-eyes are then corrected by modifying chroma, hue angles and luminance of the associated pixels such that red color is removed while maintaining a natural look of the eye. Simulation results show that the proposed algorithm is pretty robust and effective.

1 INTRODUCTION

Red-eye is a common problem in digital photography. When an image is captured with flash illumination by a camera that has an illumination source very close to the camera lens, the bright flash light reflected from the blood vessels on the retina, giving the human eyes in the image an unnatural red hue. Red-eye is a hardly acceptable defect that significantly reduces the value of an image.

For this reason, many efforts have been made to prevent it from occurring, or to detect and correct it in the post-capture processing. A straightforward way of preventing red-eye is to increase the distance between the illumination source and the camera lens. Another solution for red-eye prevention is the use of a pre-exposure flash that decreases the size of the subject's pupil followed by a second flash for capturing the image. The drawback of this approach is the great consumption of power that shortens the battery life. Moreover, this approach sometimes can only reduce, but not eliminate, the red-eye artifacts. Many research results have been developed and implemented as software products such as "Picture Maker" from Eastman Kodak Company and "iPhoto" from Apple. These products require manual manipulation to outline the red-eye region for correction. It is obviously impractical and inefficient to process a large number of images manually. A fully automatic red-eye detection and correction algorithm is therefore needed. Furthermore, it is highly expected that this automatic algorithm can be realized as a piece of hardware and planted in digital cameras. Recently, a number of researches on automatic red-eye detection and correction have been conducted (Schildkraut and Gray, 2002)-(Zhang, Sun, Li, and Zhang, 2004). The AREA algorithm proposed by Eastman Kodak Company uses features based on red-eye defects to automatically detect only a pair of red-eyes in each image (Schildkraut and Gray, 2002). In (Matthew and Robert, 2002), the face must be successfully detected before the red-eye detection where the information of color, intensity and dimension is utilized. In (Ioffe, 2003), a learning-based face detector is also adopted for the detection of red-eye defects. In these approaches, the face detection itself is another challenging problem to be solved (Xin, Xu, and Du, 1998). In (Zhang, Sun, Li, and Zhang, 2004), a heuristic algorithm is used to detect candidate red-eye

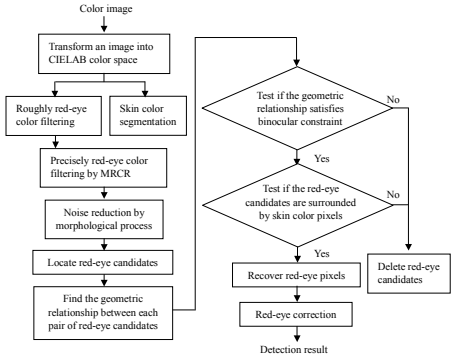


Figure 1: The function block diagram of the proposed red-eye detection algorithm.

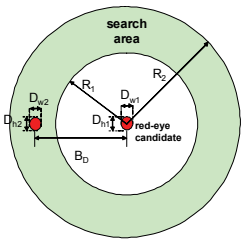


Figure 2: Binocular geometric relationship between a pair of red-eyes.

regions, and then an eye classifier is utilized to confirm whether the candidate region is a red-eye.

Color is important information to the detection and correction of red-eye artifacts. In this paper, color classifiers based on uniform color metric are first designed to detect colors of red-eyes, skin tones and colors without red hue. The relationship among these colors and the geometric constraints inherent in pairs of human eyes are exploited to eliminate false positives. Finally, the information of hue, chroma and luminance is utilized to restore the red-eye color to a natural tone.

2 UNIFORMITY OF THE CIELAB COLOR SPACE

Color is a visual perception of the light in the visible region of the electromagnetic wave spectrum incident on the human retina. By the theory of trichromacy, any color in a color space can be represented by a triple of numbers called tristimulus values (CIE, 1986), (Sangwine and Horne, 1998). However, colors in many color spaces, such as RGB, XYZ, YUV, and YC_bC_r, are not uniformly distributed in a sense that the same perceptual color difference does not correspond to the same distance enumerated in the tristimulus space (Sangwine and Horne, 1998), (Sharma and Trusell, 1997). If a color space is perceptually uniform, the perceptual difference between any two colors can be ideally represented as the Euclidean distance between their coordinates. The CIELAB color space is such a color space to overcome the non-uniform color metric that had been discussed by MacAdam (MacAdam, 1943). In this paper, the color transformation to CIELAB color space is

3 THE PROPOSED RED-EYE DETECTION ALGORITHM

The functional block diagram of the proposed red-eye detection algorithm is shown in Figure 1. The color image is first transformed to CIELAB color space. Through the uniformity in the CIELAB, the red-eye color and skin color are roughly extracted. According to the database containing all kinds of red-eye samples, a more precise red-eye color filter is then developed by the method of K-mean clustering and just-noticeable color difference in the CIELAB color space. The correlation between the size of the human eyeball and binocular distance is finally employed to eliminate most false positives

Through statistically analyzing the manually extracted red-eye color pixel and skin color pixel from various color images, the distribution of red-eye colors and skin colors in CIELAB color space can be found. For each pixel in the color image, the tristimulus values of red-eye color pixels in CIELAB color space satisfy

$$\begin{cases} L > 30 \\ 20 < a < 80 \\ -25 < b < 50 \end{cases}, \quad (1)$$

and the tristimulus values of skin color pixels in CIELAB color space satisfy

$$\begin{cases} L > 30 \\ -5 < a < 20 \\ -5 < b < 25 \end{cases}. \quad (2)$$

In this paper, the skin color is the reference information that is used to delete image regions that look like red-eyes but are not.

The method of K-means clustering is used to calculate the centroid in the red-eye set as the major red-eye color. Clustering in pattern recognition is the process of partitioning a set of pattern vectors into subsets called clusters. In this paper, the number of centroids found by the clustering method is 32. As described in Section 2, the perceptual difference

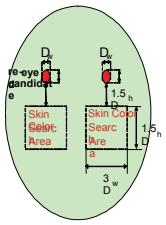


Figure 3: The surrounding skin color pixels of red-eye candidates.

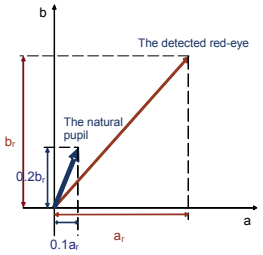


Figure 4: The color relationship between red-eye and its natural appearance on color-opponent ab plane.

between any two colors can be ideally represented as the Euclidean distance between their coordinates in the CIELAB color space, and loci of colors which are perceptually indistinguishable from a particular color form a sphere centralized at this color's coordinate. In this color space, two colors are considered perceptually distinguishable if the Euclidean distance between these two colors' coordinates exceeds a threshold of just-noticeable color difference (JNCD). That is

$$\Delta E = \left(\Delta L^2 + \Delta a^2 + \Delta b^2\right)^{1/2} \geq JNCD_{\text{Lab}} \qquad (3)$$

where $JNCD_{\text{Lab}}$ has been found around 3.0. Therefore, the major red-eye color region (MRCR) can be defined as a sphere for each major red-eye color. The colors locating in the MRCR of each centroid are selected as the color pixel in the red-eye.

In previous detection steps, some noises occur in the preliminary red-eye segmentation area. To reduce the noise, morphological process is utilized. In this paper, two morphology operations, including dilation "D" and erosion "E", are used. The former operation adds pixels to the object boundaries, while the later operation removes pixels on object boundaries in an image. The two operations are combined to build a higher order opening operation "O" for removing noises from the red-eye segmentation area while preserving the shape and size of this red-eye segmentation area in the image.

$$O(I, B_E, B_D)=D(E(I,B_E), B_D) \qquad (4)$$

where I is the binary image that marks the location of selected red-eye pixel, B_E the structuring element for erosion, and B_D the structuring element for dilation. The structuring element of dilation is smaller than that of erosion such that the morphological process can reduce noises that occur in the red-eye segmentation area and avoid losing the red-eye candidates.

By using color information and morphological process, red-eye candidates are located. However, many detected image regions that look like red-eyes but are not. The color pixels in such regions are called false positive candidates. Hence, geometric relationship between each pair of red-eye candidates is considered. Since the distance between a pair of red-eyes is useful to eliminate the false positive candidate, binocular geometric relationship is used to develop binocular geometric constraint for removing false positives. For a red-eye candidate shown in Figure 2, the green circular region centralized at this candidate is used to test whether its red-eye counterpart locates in this region. From the statistical analysis of a large number of data sets, the range of the green circular region is constrained by

$$R_1 > 4D_{w1}, \qquad (5)$$

$$R_2 < 10D_{w1}. \qquad (6)$$

The pair of red-eye candidates is further tested if its width and height can satisfy

$$0.75 < D_{w2}/D_{w1} < 1.30, \qquad (7)$$

$$0.75 < D_{h2}/D_{h1} < 1.30. \qquad (8)$$

If the width and height of the pair of red-eye candidates cannot satisfy Eq. (7) and (8), the pair of red-eye candidates will be removed.

Finally, the pair of red-eye candidates that pass through the binocular geometric constraint is tested if it is surrounded by pixels of skin color tones. The surrounding skin color pixels of red-eye candidates are simply defined as two areas as shown in Figure 3. The width and height of the area are $1.5D_h$ and $3D_w$, respectively. The area is recognized as skin color if the amount of skin color pixels, N_{skin}, that satisfy Eq. (2) is high enough, or

$$N_{skin}/(1.5D_h \times 3D_w) > 0.8. \qquad (9)$$

4 RECOVERY AND COLOR CORRECTION OF RED-EYES

Since some red-eye color pixels are removed from the red-eye candidates in the previous processes, it is

therefore required to retrieve the lost red-eye color pixels. The region of the red-eye candidate is simply extended by its neighboring red-eye color pixels. That is, the boundary of the red-eye candidate region centered on its center location is enlarged pixel by one pixel to form a perfectly retrieved region of the candidate.

Once the location and size of the red-eyes have been detected, color correction of red-eyes is applied to the detected red-eyes to obtain the natural appearance of the pupil. To maintain the natural appearance of the pupil at the location of red-eyes, the value of luminance (L component) of the detected red-eye is slightly adjusted and the values of hue and chroma (a and b components) are adjusted based the color relationship between red-eye and the corresponding natural appearance of pupils on color-opponent ab plane. For simplicity, the adjustment for color correction is to scale down the value of a component of red-eyes by a factor of 0.1 and to scale down the value of b component of red-eyes by a factor of 0.2 as shown in Figure 4. The value of luminance of the detected red-eye is adjusted by a factor of 0.9. That is,

$$L_{corrected} = 0.9 \times L_r \quad (10)$$

$$a_{corrected} = 0.1 \times a_r \quad (11)$$

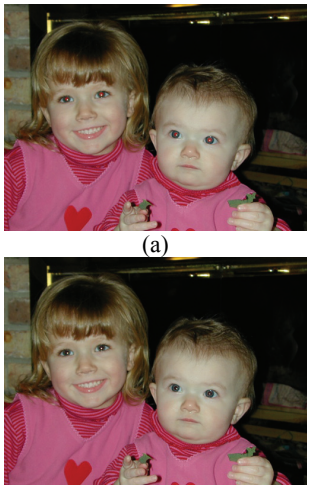
$$b_{corrected} = 0.2 \times b_r \quad (12)$$

where (L_r, a_r, b_r) and $(L_{corrected}, a_{corrected}, b_{corrected})$ are tristimulus values of the detected red-eye color pixel and its corrected color pixel, respectively.

5 SIMULATION RESULTS AND CONCLUSIONS

To evaluate the performance of the proposed algorithm, the simulation of the red-eye detection algorithm that is applied to red-eye digital images with different size and quality is conducted. In Figure 5, the "Pinksisters" image that has more than one pairs of red-eyes is also detected and corrected by using the proposed algorithm. In our experiments, over 200 red-eye digital photographs are tested and more than 80% red-eyes are efficiently detected. The experimental results show that the proposed algorithm is robust and effective under a variety of shooting conditions and backgrounds.

In this paper, a fully automatic red-eyes detection and correction algorithm is proposed. In the proposed algorithm, a robust color classifier for detecting red-eye color and other major colors in digital images with red-eyes is developed and a multi-stage criterion for detecting each single red-eye is designed. The detected red-eyes are successfully corrected by modifying chroma, hue angles and luminance of the associated pixels such that red color is removed while maintaining a natural look of the eye. The proposed system has very low false detection rate. Simulation results show that more than 80% of red-eyes can be detected and only 5% are false alarm.



(a)

Figure 5 (a): "Pinksisters" image with red-eyes, (b) the image after correcting red-eye colors.

REFERENCES

Schildkraut, J.S. and Gray, R.T., 2002. "A fully automatic red-eye detection and correction algorithm," in *Proc. Int. Conf. on Image Processing ICIP*, pp. I-801~803.

Matthew, G. and Robert, U. 2002. "Automatic red-eye detection and correction," in *Proc. Int. Conf. on Image Processing ICIP*, pp. I-804~807.

Ioffe, S. 2003. "Red eye detection with machine learning," in *Proc. Int. Conf. on Image Processing ICIP*, pp. II-871~874.

Zhang, L., Sun, Y. F., Li, M. J., and Zhang, H. J., 2004. "Automated red-eye detection in digital photography," in *Proc. Int. Conf. on Image Processing ICIP*, pp. 2363-2366.

Xin, Z., Xu, Y. J., and Du, L. M., 1998. "Locating facial features with color information," in *Proc. Int. Conf. on Image Processing ICIP*, Vol. 2, pp. 889-892.

CIE, 1986. *Colorimetry*, CIE Pub. No. 15.2, CIE Central. Bureau, Vienna, Austria.

Sangwine, S. J. and Horne, R. E. N., 1998. *The Colour Image Processing Handbook*. London: Chapman and Hall.

Sharma, G. and Trusell, H. J. 1997. "Digital color imaging," *IEEE Trans. on Image Processing*, vol. 6, pp. 901-932.

MacAdam, D. L., 1943. "Specification of small chromaticity differences," *J. Opt. Soc. Am*, vol. 33, pp. 18-26.

SPEAKER RECOGNITION USING DECISION FUSION

M. Chenafa, D. Istrate
RMSE, ESIGETEL, 1 Rue du Port de Valvins, 77215 Avon-Fontainebleau, France
mohamed.chenafa@esigetel.fr, dan.istrate@esigetel.fr

V. Vrabie, M. Herbin
CReSTIC, Université de Reims Champagne-Ardenne, Chaussée du Port, 51000 Châlons-en-Champagne, France
valeriu.vrabie@univ-reims.fr, michel.herbin@univ-reims.fr

Keywords: Biometrics, Speaker recognition, Speech recognition, Decision fusion, GMM/UBM.

Abstract: Biometrics systems have gained in popularity for the automatic identification of persons. The use of the voice as a biometric characteristic offers advantages such as: is well accepted, it works with regular microphones, the hardware costs are reduced, etc. However, the performance of a voice-based biometric system easily degrades in the presence of a mismatch between training and testing conditions due to different factors. This paper presents a new speaker recognition system based on decision fusion. The fusion is based on two identification systems: a speaker identification system (text-independent) and a keywords identification system (speaker-independent). These systems calculate the likelihood ratios between the model of a test signal and the different models of the database. The fusion uses these results to identify the couple speaker/password corresponding to the test signal. A verification system is then applied on a second test signal in order to confirm or infirm the identification. The fusion step improves the false rejection rate (FRR) from $21,43\%$ to $7,14\%$ but increase also the false acceptation rate (FAR) from $21,43\%$ to $28,57\%$. The verification step makes however a significant improvement on the FAR (from $28,57\%$ to 14.28%) while it keeps constant the FRR (to $7,14\%$).

1 INTRODUCTION

Biometric recognition systems, which identify a person on his/her physical or behavioral characteristics (voice, fingerprints, face, iris, etc.), have gained in popularity among researchers in signal processing during recent years. Biometric systems are also useful in forensic work (where the task is whether a given biometric sample belongs to a given suspect) and law enforcement applications (Atkins, 2001). The use of the voice as a biometric characteristic offers the advantage to be well accepted by users whatever his culture. There are two categories in voice-based biometric systems: speaker verification and speaker identification. In identification systems, an unknown speaker is compared to the N known speakers stored in the database and the best matching speaker is returned as the recognition decision. Whereas in verification systems, an identity is claimed by a speaker, so the system compares the voice sample to the claimed speaker's voice template. If the similarity exceeds a predefined threshold, the speaker is accepted, otherwise is rejected. For each system two methods can be distinguished: text-dependent and text-independent. In the first case, the text pronounced by the speaker is known beforehand by the system, while in the second case the system does not have any information on the pronounced text (Kinnunen, 2003).

It is well known that the performances of voice-based biometric systems easily degrade in the presence of a mismatch between the training and testing conditions (channel distortions, ambient noise, etc.). One method that can be used to improve the performances of these systems is to merge various information carried by the speech signal. Several studies on information fusion were led to improve the performances of automatic speakers recognition system (Higgins et al., 2001)(Mami, 2003)(Kinnunen et al., 2004). However, the results are less successful compared to biometric systems based on other modalities (fingerprint, iris, face, etc).

In this paper a new fusion approach is proposed by using two kinds of information contained in the speech signal: the speaker (who spoke ?) and the keyword pronounced (what was said ?). The aim of this method is to use a first test signal to identify a couple speaker/password corresponding to this signal. This step is done by combining two identification systems based on likelihood ratio approach: a speaker identification system (text-independent) and a speech iden-

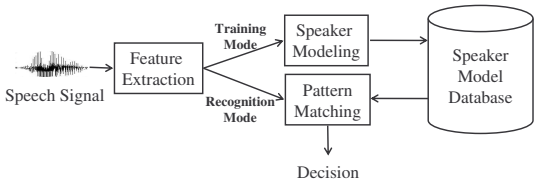


Figure 1: Components of a speaker recognition system.

tification system (speaker-independent). The speaker identified by this fusion is then verified by a classical verification text dependent system using a second test signal. In practical situations, the two test signals can be viewed as a composed password. The system provides good improvements on the two types of error usually computed for biometric systems: the false rejection rate (FRR) and the false acceptation rate (FAR). The experiments presented in this study use the platform ALIZE developed by the LIA laboratory (Bonastre et al., 2005).

This paper is organized as follows. Section 2 provides a general description of a speaker recognition system. Section 3 presents the proposed fusion system. The experiments are discussed in Section 4, followed by conclusions in the last section.

2 SPEAKER RECOGNITION SYSTEM

Figure 1 shows the structure of an automatic speaker recognition system. This system operates in two modes (training and recognition) and can be used for both identification or verification tasks. In the training mode, a new speaker (with known identity) is enrolled into the system's database, while in the recognition mode an unknown speaker gives a speech input and the system makes a decision about the speaker identity.

2.1 Feature Extraction

Feature extraction is the first component in an automatic speaker recognition system (Furui, 1997). This phase consists of transforming the speech signal in a set of feature vectors called also parameters. The aim of this transformation is to obtain a new representation which is more compact, less redundant, and more suitable for statistical modeling and calculation of distances. Most of the speech parameterizations used in speaker recognition systems relies on a cepstral representation of the speech signal (Lee et al., 1996).

2.1.1 MFCC and LFCC Parameters

The *Mel-frequency cepstral coefficients* (MFCC) are motivated by studies of the human peripheral auditory system. Firstly, the speech signal $x(n)$ is divided into Q short time windows which are converted into the spectral domain by a Discret Fourier Transform(DFT). The magnitude spectrum of each time window is then smoothed by a bank of triangular bandpass filters (Figure 2) that emulate the critical band processing of the human ear.

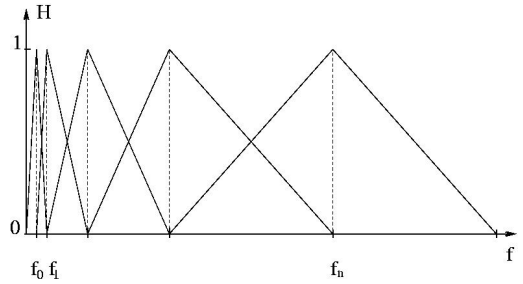


Figure 2: Mel filter bank.

Each one of the bandpass filter $H(k,m)$ computes a weighted average of that subband, which is then logarithmically compressed:

$$X'(m) = \ln\left(\sum_{k=0}^{N-1} |X(k)|\, H(k,m)\right) \qquad (1)$$

where $X(k)$ is the DFT of a time window of the signal $x(n)$ having the length N, the index k, $k = 0,\ldots,N-1$, corresponds to the frequency $f_k = kf_s/N$, with f_s the sampling frequency, the index m, $m = 1,\ldots M$ and $M << N$, is the filter number, and the filters $H(k,m)$ are triangular filters defined by the center frequencies $f_c(m)$ (Sigurdsson et al., 2006). The log compressed filter outputs $X'(m)$ are then decorrelated by using the Discrete Cosine Transform (DCT):

$$c(l) = \sum_{m=1}^{M} X'(m)\cos(l\frac{\pi}{M}(m-\frac{1}{2})) \qquad (2)$$

where $c(l)$ is the l^{th} MFCC of the considered time window. A schematic representation of this procedure is given in Figure 3.

There are several analytic formulae for the Mel scale used to compute the center frequencies $f_c(m)$. In this study we use the following common mapping:

$$B(f) = 2595\log_{10}(1+\frac{f}{700}) \qquad (3)$$

The LFCC parameters are calculated in the same way as the MFCC, but the triangular filters use a linear frequency repartition.

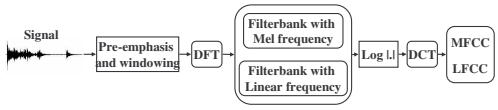


Figure 3: Extraction of MFCC and LFCC parameters.

2.1.2 Δ and $\Delta\Delta$ Parameters

After the cepstral coefficients have been calculated and stored in vectors, a dynamic information about the way these vectors vary in time is incorporate. This is classically done by using the Δ and $\Delta\Delta$ parameters, which are polynomial approximations of the first and second derivatives of each vector (Kinnunen et al., 2004).

2.2 Speaker Modeling

The training phase uses the acoustic vectors extracted from each segment of the signal to create a speaker model which will be stored in a database. In automatic speaker recognition, there are two types of methods that give the best results of recognition: the deterministic methods (dynamic comparison and vector quantization) and statistical methods (Gaussian Mixture Model - GMM, Hidden Markov Model - HMM), these last ones being the most used in this domain. In this paper, we have chosen to use a system based on GMM-UBM. This choice was motivated by two reasons: modeling by GMM is very flexible with regard to the type of the signal and using the GMM gives a good compromise between performances and the complexity of the system.

2.2.1 GMM-UBM

In this research, the method used for speaker modeling is the GMM using the universal background model (UBM). The UBM has been introduced and successfully applied by (Reynolds, 1995) in speaker verification. This model is created by using all recording speech of the database, the aim being to have a general model of speakers which will be then used to adapt each speaker model.

The matching function in GMM is defined in terms of the log likelihood of the GMM (Bimbot et al., 2004) given by:

$$p(X|\lambda) = \sum_{q=1}^{Q} log\ p(x_q|\lambda) \tag{4}$$

where $p(x_q|\lambda)$ is the Gaussian mixture density of the q^{th} segment in respect to the speaker λ:

$$p(x_q|\lambda) = \sum_{i=1}^{G} p_i f(x_q|\mu_i^{(\lambda)}, \Sigma_i) \tag{5}$$

with the mixing weights constrained by:

$$\sum_{i=1}^{G} p_i = 1 \tag{6}$$

In these expressions x_q is the D-dimensional acoustic vector corresponding to the q^{th} time window of the input signal, p_i, $\mu_i^{(\lambda)}$ and Σ_i $(i = 1,\ldots,G)$ are the mixture weight, mean vector, and covariance matrix of the i^{th} Gaussian density function (denoted by f) of the speaker λ, while G denotes the number of GMM used by the model.

The speaker model λ is thus given by:

$$\lambda = \left\{ p_i, \mu_i^{(\lambda)}, \Sigma_i | i = 1,\ldots,G) \right\} \tag{7}$$

the UBM model having the same form:

$$UBM = \left\{ p_i, \mu_i^{(UBM)}, \Sigma_i | i = 1,\ldots,G \right\} \tag{8}$$

The mean vectors of speaker model $\mu_i^{(\lambda)}$ are adapted to the training data of the given speaker from the UBM, i.e. $\mu_i^{(UBM)}$, by using the Maximum a Posteriori (MAP) adaptation method (Gauvain and Lee, 1994), the covariance matrices and mixture weights remaining unchanged.

2.3 Pattern Matching and Decision

Given a segment of speech, Y, and a hypothesized speaker, S, the task of speaker recognition system is to determine if Y was spoken by S. This task can be defined as a basic hypothesis test between

$$\begin{cases} H_0\text{: Y is from the hypothesized speaker S} \\ H_0\text{: Y is not from the hypothesized speaker S} \end{cases}$$

To decide between these two hypotheses, the optimum test is a likelihood ratio given by:

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \ \theta & Accept \quad H_0 \\ < \ \theta & Reject \quad H_0 \end{cases} \tag{9}$$

where $p(Y|H_i)$ is the probability density function for the hypothesis H_i evaluated for the observed speech segment Y, also referred to the likelihood of the hypothesis H_i. The decision threshold for accepting or rejecting H_0 is θ. A good technique to compute the values of the two likelihoods, $p(Y|H_0)$ and $p(Y|H_1)$ is given in (Doddington, 1985).

3 PROPOSED SYSTEM

In this paper a new method for automatic speaker recognition based on fusion information is proposed. The architecture of this method is described in Fig. 4.

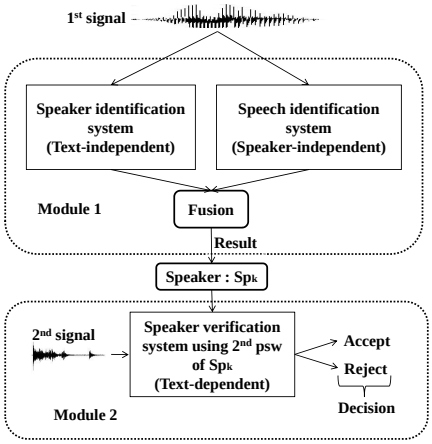


Figure 4: Global system architecture.

This system is composed by two blocks, the first one made up by two classifiers (speaker and password classifiers) and the second one made up by a verification system using the decision result of the first block. Each speaker is identified by two passwords: the first one is used by both speaker and password identification systems, while the second one by the verification system. In practical situations, these two passwords can be viewed as a composed password.

The identification systems (speakers and passwords identification) are used in open-set (no information available on the possible speakers and passwords). Both systems calculate the likelihood ratio on a first test signal by using equation (4). We used here a normalization model UBM, as presented in section 2.2.1. This means that during the creation of the models (speaker, password), each model is adapted by the MAP method to the UBM model.

The verification system is a classical speaker verification system which is used to confirm or infirm the speaker identified previously by using a second test signal (the second password).

Figure 5 shows the fusion between speaker and speech (password) identification systems.

After sorting the *log* likelihood ratios (for the first test signal) calculated with regard to the speakers model $LLK(X|Sp_i)$, $i = 1,N$ (N is the number of speakers stored in the database) and to the passwords model $LLK(X|Psw_i)$,$i = 1,N$ (N is the number of passwords stored in the database), a first test consists of comparing the most likely speaker given by the speaker classifier with the first three identified passwords given by the password classifier. If his password was found between the three identified passwords, a couple (speaker/password) was thus identified. A second test consists of comparing the most likely password with the first three identified speakers. If this password belongs to one of them, another couple (password/speaker) is identified. In the cases where two couples are identified, the couple with the biggest likelihood ratio (Lk_Sp + Lk_P) is retained. The system can reject directly a recording if there are no identified couples.

Once the first test signal is associated to a speaker, a classical verification is then launched using the second test signal pronounced by the speaker identified previously. If the likelihood ratio of this verification is smaller than the smallest likelihood ratio of the first two recordings used in the training phase, the identity of the speaker is confirmed, otherwise the speaker is rejected.

4 EXPERIMENTS

4.1 Data Base

In order to evaluate the proposed system a corpus of specific keywords has been recorded. This corpus contains the recordings of 15 isolated words (French language) and 11 numbers (from 0 to 10).
The recordings were stored in WAV format, with a sampling rate $f_s = 16$ kHz. The parameterization was realized by using MFCC parameters for the passwords identification system and LFCC for speaker identification and verification systems. We have optimized the acoustic parameter for this application; all the 8 ms the signal is characterized by a vector made up of 16 ceptrals coefficients $c(l)$ (see Eq. (2)) and their derivative $\Delta\Delta$.

4.2 Training and Test Data

For both identification systems (speaker and password) the first password recording is used for the

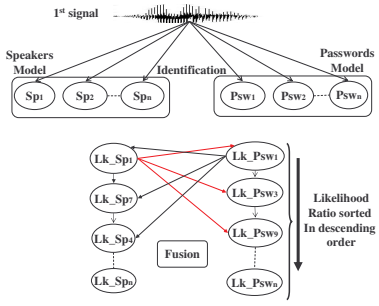


Figure 5: Fusion system architecture(Module 1).

training mode. The verification system uses the second password for the training mode. The speakers database is divided into two equals groups: 7 clients and 7 impostors. Therefore, in the test stage the number positive and negative tests are equals.

1. The speaker identification system (text-independent) uses two recordings of 14 words of the 7 clients for the training phase. For the recognition phase, the system uses one recording of 14 words of the 7 clients and 3 recordings of 14 words of 7 impostors.
2. The password identification system (speaker-independent) uses two recordings of 7 clients for the training phase. For the recognition phase, the system uses one recording of 7 clients and all recordings of the impostors.
3. The verification system uses two recordings of the second passwords of every client for the training phase and a recording of 7 clients as well as all the recordings of 7 impostors for the recognition phase.
4. The reference system uses for the training phase 7 speakers, two recordings of 14 words. For recognition phase we used a recording of 14 speakers passwords and 3 recordings of 14 impostors words.

4.3 Reference System

The results obtained by the global system are compared to a classical verification system (Bimbot et al., 2004). In the training phase of the reference system a speaker model is created from the feature vectors (16 LFCC + $\Delta\Delta$) using two recordings of all the passwords to model speakers; However the recognition phase uses all passwords of the speakers pronounced by impostor and other words.
We have optimized the number of GMM for this application; the optimal value is $G = 16$.

4.4 Results and Discussion

Table 1 shows the false rejection rate (FRR) and the false acceptation rate (FAR) of the reference system, the first module of the new system and the global system proposed.

The best equal error rate obtained for the reference system is 21.43%, which is high enough but can be justified by the small size of the database. After the fusion of the results between the speaker identification system and the password identification system, we notice that the FAR increases to 28.57% (that is due to the password identification system which increases the chance of impostors to be accepted because the password is well recognized), while the FRR decreases to 7.14%. By using a verification system, which uses the results of this fusion, we improve the FAR (from 28.57% to 14.28%) while the FRR remains the same one (7.14%) because the verification system was adapted to recognize the clients. The global system thus makes an improvement of 43.47% of the FRR and 65.69% of the FAR. Note again that these values are high enough due to the small size of the database.

Table 1: performances of different systems.

Systems	FRR	FAR
Reference System text dependent (16 LFCC + $\Delta\Delta$)	21.43%	21.43%
Fusion System between speakers identification (16 LFCC+$\Delta\Delta$) and passwords identification (16 MFCC + $\Delta\Delta$)	7.14%	28.57%
Verification after fusion (16 LFCC + $\Delta\Delta$)	7.14%	14.28%

5 CONCLUSIONS

In this paper, we presented several experiments to improve the performances of a voice-based biometric system using decision fusion. The fusion of the speaker identification and the passwords identification was firstly proposed. We show that the fact of modeling the passwords pronounced by the speakers brings improvements in the false reject rate but in the same time it increases the number of the impostors accepted by the system. The second experience proposes an automatic speaker verification using the result (speaker identified) of the first experience. The aim here is to confirm the results returned by the fusion of speaker and password classifiers. This second experience allows us to reduce the number of impostors accepted by the system and improves the results of the fusion by decreasing the FAR from 28.57% to 14.28%. So the global system improves the performances in term of FAR and FRR with regard to the reference system. This study encourages us to continue the experimentation on a corpus with more important size and to consider other kind of fusion such as weigthed ranks.

REFERENCES

Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Chagnolleau, I., Meignier, S., Merlin, T., Garciya, J., Delacrtaz, D., and Reynolds, D. (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 4:430–451.

Bonastre, J.-F., Wils, F., and Meignier, S. (2005). Alize, a free toolkit for speaker recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 737–740.

Doddington, G. (1985). Speaker recognition - identifying people by their voices. In *Proc. of the IEEE*, volume 73, pages 1651–1664.

Furui, S. (1997). Recent advances in speaker recognition. In *Proc. of the First International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 1206, pages 237–252.

Gauvain, J.-L. and Lee, C.-H. (1994). Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. In *IEEE Trans. on Speech and Audio*, volume 2, pages 291–298.

Higgins, J. E., Damper, R. I., and Harris, C. J. (2001). Information fusion for subband-hmm speaker recognition. In *International Joint Conference on Neural Networks*, volume 2, pages 1504–1509.

Kinnunen, T. (2003). *Spectral Features for Automatic Text-Independent Speaker Recognition*. PhD thesis, University of Joensuu, Finland.

Kinnunen, T., Hautamki, V., and Fr´anti, P. (2004). Fusion of spectral feature sets for accurate speaker identification. In *9th International Conference Speech and Computer (SPECOM)*, pages 361–365.

Lee, C. H., Soong, F., and Paliwal, K. (1996). *Automatic Speech and Speaker Recognition*. Springer, London, UK, 2nd edition edition.

Mami, Y. (2003). *Reconnaissance de locuteurs par localisation dans un espace de locuteur de reference*. PhD thesis, ENST Paris, France.

Reynolds, D. (1995). Speaker identification and verification using gaussian mixture speaker models. *Speech Communication*, 17:91–108.

Sigurdsson, S., Petersen, K. B., and Lehn-Schiøler, T. (2006). Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music. In *Proceedings of the Seventh International Conference on Music Information Retrieval (ISMIR)*, pages 286–289.

SOBI WITH ROBUST ORTHOGONALIZATION TO REMOVE THE ARTEFACT STIMULUS IN EVOKED POTENTIAL

5Hz Current Sinusoidal Stimulus

Eduardo de Queiroz Braga, Carlos Julio Tierra-Criollo
Department of Electrical Engineering, UFMG,Av. Presidente Antônio Carlos 6627,Belo Horizonte,Brasil
eduardo@cpdee.ufmg.br ,carjulio@cpdee.ufmg.br

Gilberto Mastrocola Manzano
Clinical Neurophysiology Laboratory , UNIFESP, São Paulo, Brasil
gmanzano@uol.com.br

Keywords: SOBI-RO, somatosensory system, artefacts.

Abstract: The psychophysical evaluation of the sensibility of the thin and thick fibers with sinusoidal current stimulation was proposed in the 80s. After that, researches observed that 5 Hz stimulus would be related to the thin unmyelinated fiber. This work aims a quantitative analysis of the cerebral cortex response to 5 Hz stimulus, through the identification of the latency components of the evoked potential (EP) that were estimated by the coherent mean after remove the stimulus artefact by using the Independent Component Analysis. Electroencephalography (EEG) signals were collected at Cz electrode (10-20 International Standard System) of 5 volunteers. The EP estimated with 5 Hz stimulus using the Second Order Blind Identification associated with Robust Orthogonalization (SOBI-RO) associated with the coherent mean presented the following components: N_1 = 104 ms (one volunteer), P_1 = 179 ms (four volunteers) and N_2 = 234 ms (three volunteers), P_2 = 280 ms (three volunteers) and N_3 = 493 ms (all volunteers). The SOBI-RO techniques can be a very useful tool in artefacts and noise reduction on the EP estimation.

1 INTRODUCTION

Our knowledge about the world is built over different sensations. The perceptions begin at receptors cells and are transmitted to the central nervous system through primary afferents fibers .In the somatic system, these fibers have different diameters and transmit different sensations to the spinal cord: thin fibers transmit pain and temperature, and thick fibers transmit the sense of touch. An instrument of psychophysical sensibility evaluation, proposed in the 80's, is based on the principle that activation of different diameters fibers depends on frequency of sinusoidal currents: 5 Hz to non-myelinic fibers (Masson et al., 1989; Ro et al., 1989), 250 Hz to thin myelinic fibers and 2 kHz to thick myelinic fibers.

The evoked potential (EP) by electric stimulus can be obtained using the coherent mean (Misulis, 1994; Regan, 1989). When a sinusoidal current of 5 Hz is used to stimulate, a strong level of artefact in this frequency is collected in the EEG electrodes. The 5 Hz artefact damage the EP and the extraction of this artefact (synchronised to the stimulus) is very difficult because of is into EP frequencies. In this case, alternative tools can be used. In this context, the use of statistics tools can help us. The Second Order Blind Identification associated with Robust Orthogonalization -SOBI-RO (Belouchrani et al., 1997; Belouchrani and Cichocki, 2000) can be a useful technique where the stimulus artefact is presented in the same frequency band of the EP. It can be applied in EEG electrodes that are spatially located in the scalp where each electrode is considered like a linear mixture of blind brain sources.

In the present work, the SOBI-RO was used to detect and remove independent components associated with the artefact and rhythm that difficult the analysis on Cz channel. The reconstructed signals would present the epochs without the artefacts, and then, the ERP could be better identified using the coherent mean.

2 MATERIALS AND METHODS

The EEG signals were collected in 5 normal volunteers with closed eyes, without neurological disease or medication. The experimental protocol was performed in the Clinical Neurophysiology laboratory on UNIFESP and was approved by the Local Ethic's Committee. The electrodes of stimulation (10mm diameter gold electrodes) were placed in the medial and lateral surfaces of the distal phalanx, of the second finger of the left hand with a thin amount of conductive gel. The 5 Hz sinusoidal current stimulus with twice the sensibility threshold was applied by the Neurometer Current Perception Threshold (CPT)-USA. The Electroencephalogram (EEG) signals were collected in the Cz channel and the reference was A1+A2 (connected ear). In addition, the stimulus signals were collected on left wristband (Pi). These signals were used for synchronization of the epochs. Six sessions with one hundred of epochs (20s each, followed by 10s without stimulation) were recorded with a sample rate of 500 Hz by the NeuroScan SymAmpsTM – USA. In each epoch were extracted two seconds before and six after the stimulus where it expects to find the EP. The 100 epochs of 8 seconds were applied in the SOBI-RO algorithm labelled *ICALAB* 2.5 for *MATLAB* (ICALAB 2004).

2.1 The SOBI-RO

The SOBI-RO (*Second Order Blind Identification with Robust Orthogonalization*) is a statistic tool of ICA (*Independent Component Analysis*). This tool considers the measured signals like a linear combination of unknown sources (Hyvrinen et al, 2001). In this context, the epochs **x** can be expressed like:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + ... + a_{1n}s_n(t) \\ x_2(t) &= a_{21}s_1(t) + ... + a_{2n}s_n(t) \\ &\vdots \\ x_m(t) &= a_{m1}s_1(t) + ... + a_{mn}s_n(t) \end{aligned} \quad (1)$$

Or can be represented as:

$$\mathbf{x} = \mathbf{A}.\mathbf{s} \quad (2)$$

Where **X** is the epochs collected in Cz channel and synchronized whit the stimulus. **A** is an unknown mixing matrix that make the data **x** a linear combination of the unknown sources **s.**

A pre step in the ICA is the *Whitening*. It is used to represent the data in a new space, where the signals are decorrelated with exhaustion. Belorachrin and Cichocki (2000) presented a robust technique applied in the whitening process called Robust Orthogonalization that can give us a better estimation of the coefficients of the whitening matrix **W**.

In the Robust Whitening, a set of covariance matrices of **x** at different lags is used to estimate the *whitening* matrix:

$$\mathbf{R}_x(\tau) = \mathrm{E}[\mathrm{x(t)}.\mathrm{x}^*(\mathrm{t}-\tau)] = \mathbf{A}\mathbf{R}_S(\tau)\mathbf{A}^H \quad (3)$$

Where $\tau=1,...,K$

The method uses an optimization algorithm that estimate a linear combination of evaluated covariance's matrices R_X:

$$\mathbf{C} = \sum_{\tau=1}^{K} \alpha_\tau \hat{\mathbf{R}}_X(\tau) \quad (4)$$

The eigen value decomposition (EVD) of C is performed:

$$\mathbf{C} = \mathbf{U}_C diag[\lambda_1^2,...,\lambda_n^2]\mathbf{U}_C^T \quad (5)$$

And the whitening matrix is:

$$\mathbf{Z} = diag[\lambda_1,...,\lambda_n]^{-1}\mathbf{U}_C^T \quad (6)$$

The whitened data **z** is expressed like:

$$\mathbf{Z} = \mathbf{W}.\mathbf{x} = \mathbf{W}.\mathbf{A}.\mathbf{x} \quad (7)$$

W.A is a unitary matrix **U**. In this context, the objective of SOBI is to discover this matrix **U**. For this, a cost function called *join diagonalyzer* -JD (Belouchrani et al., 1997) is used. For that, a set of covariance matrices of the data **z** is taken at different lags:

$$\hat{\mathbf{R}}_\mathbf{Z}(\tau) = \hat{\mathbf{W}}\hat{\mathbf{R}}_\mathbf{X}(\tau)\hat{\mathbf{W}}^T \quad (8)$$

Using second order information for theses matrices it is possible to find the matrix **U** by an optimization method of search.

Then, the mixing matrix A and the sources can be estimated by:

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}} \quad (9)$$

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^H\hat{\mathbf{W}}\mathbf{x}(t) \quad (10)$$

where # is a pseudo-inverse matrix and H is a Hermitian matrix.

2.2 Application of SOBI-RO

After the SOBI-RO detection, the independent components passed by a visual inspection, and the components related with the 5Hz stimulus were deselected. The new epochs were reconstructed and the coherent mean applied. But in this average, the alpha rhythm was strongly present. Thus, in a second approach, the SOBI-RO was applied to remove frequency components of 8-10Hz that can be associated with spontaneous EEG.

3 RESULTS

The EP for volunteer #1, obtained with the original EEG signal at Cz channel (Figure 1a), presents high level of the 5 Hz artefact that difficult the analysis. After removing this 5 Hz artefact with SOBI-RO, the EP can be seen most clearly in the Figure 1b. The power spectral density (PSD) shows the attenuation of the 5 Hz frequency and odd harmonics of 5 Hz (Figure 2).

A rhythm into 8-10Hz frequencies is also presented, but before and after stimulation (Figure 1.b). The new EP shows the attenuation of this band (Figure 2a and 2b). The components identified in this EP (Figure 3, Table 1) were: $P_1 = 188$ ms, N_1=234 ms, $P_2 = 268$ ms and $N_2 = 441$ ms. The grand-average of the five volunteers EP's presented components at $N_1 = 109$ ms, $P_1 = 200$ ms, $N_2 = 230$ ms, $P_2 = 279$ ms and $N_3 = 441$ ms (Table 1).

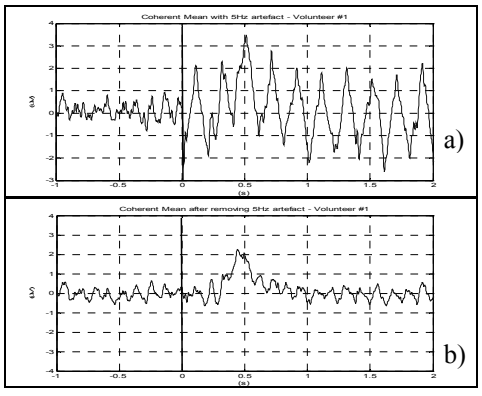


Figure 1: EP of Cz channel (volunteer #1), (a) before and (b) after SOBI-RO removing 5Hz component. Time 0 s represents the beginning of the stimulation.

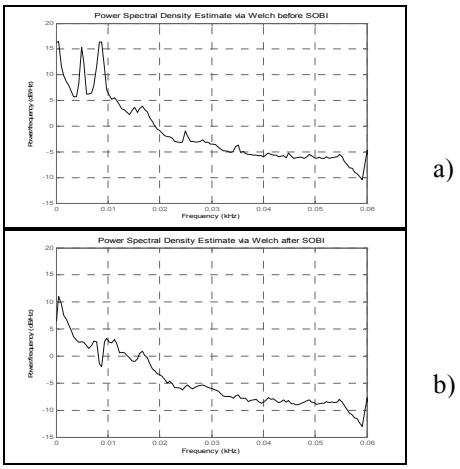


Figure 2: PSD of EEG signals (volunteer #1) (a) before and (b) after SOBI-RO.

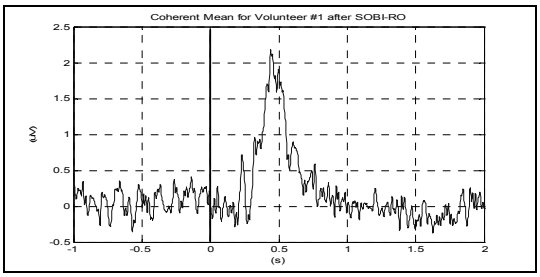


Figure 3: The EP after removing the artefact and 8-10Hz related IC's.

4 DISCUSSIONS

During the process to remove the 5Hz artifact with SOBI-RO, the IC that represents this frequency was clearly identified and removed. We can see in the PSD (Figure 2) that the 5Hz stimulus artifact and odd harmonics were completely removed. This shows that the SOBI-RO was efficient in this step. However, on the process for identifying of IC's related to the 8-10Hz band (possibly, associated with the spontaneous alpha rhythm during closed eyes) was more difficult. For each volunteer, ten or more IC's related with this band were founded. Some IC's showed a variation of the amplitude with the stimulus .This fact does doubtful their removals and suggests the necessity of a better method of detection based on the statistical information of the IC's. In this work, the procedure was repeated ten times (using 70 epochs randomly selected each time) for evaluating the experimenter bias (due the visual selection of the IC's). The results were similar in all cases.

The EP identification was only possible after SOBI-RO pre-processing. The N_1 component was only identified in the EP of volunteer #2. This component will be confirmed in future researches, with more experiments. In the other hand, the components P_1, N_2 and P_2 were present in a great number of volunteers. All volunteers presented the N_3 component (between 441ms and 604 ms). The grand average also shows the N_1, P_1, N_2, P_1 e N_3 components (Table 1).

Table 1: Components of the EP of five volunteers and Grand Average after applying SOBI-RO.

Volunteer	N_1 (ms)	P_1 (ms)	N_2 (ms)	P_2 (ms)	N_3 (ms)
#1	-	188	234	268	441
#2	104	206	-	-	424
#3	-	139	237	283	562
#4	-	181	230	290	434
#5	-	-	-	-	604
Mean	104	179	234	280	493
Standard deviation	-	28	4	11	84
Grand Average					
	109	200	230	279	441

5 CONCLUSIONS

This work presented a useful application of SOBI-RO with the objective of removing the 5 Hz sinusoidal current artefact and spontaneous activity in the 8-10 Hz band. The conventional filtering can not remove theses frequency bands without remove information of the EP.

Research with SOBI-RO can be very useful in signals where the artefact stimulus frequency is in the same band of the EP.

ACKNOWLEDGEMENTS

The author would like to thank to the FAPEMIG, FAPESP, CNPq and CAPES by the financial support.

REFERENCES

Belouchrani A., Cichocki A., 2000. *Robust Whitening Procedure in Blind Source Separation Context*, Electronics Letters, vol. 36, p. 2050-2051.

Belouchrani, A., Meraim, K.A., Cardoso, J.-F., and E. Moulines, 1997. *A blind source separation technique using second order statistics,* IEEE Trans. Signal Processing, vol. 45, pp. 434-444.

Hyvrinen, A., Karhunen, J., Oja, E. 2001. *Independent Components Analysis*, John Wiley & Sons Canada, Ltd. 1st ed. Canada

ICALAB: free software. Version 2.2. Japan, RIKEN Labs., 2004. Posted in: <http://www.bsp.brain.riken.jp/ICALAB/>. Access in July, 10, 2006.

Masson, E.A., Veves, A., Fernando, D., Boulton, A.J.M. (1989), Current perception thresholds: anew, quick, and reproducible method for the assessment of peripheral neuropathy in diabetes mellitus, Diabetologia, 32, p.724-728.

Misulis, K. E., 1994. *Spellman's Evoked Potential Primer: visual, auditory, and somatosensory evoked potentials in clinical diagnosis*. Ed. Butterworth-Heinemann, Boston.

Regan, D. 1989. *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine.* Ed. Elsevier, New York. P.48-57.

Ro, L., Chen, S., Tang, L., Hsu, W., Chang, H., Huang, C. (1999), *Current perception threshold testingin Fabry's disease*. Muscle & Nerve, 22, p. 1531-1537.

COMPUTERISED SYSTEM FOR EVALUATION OF ASYMMETRY OF POSTURAL PARAMETER COEFFICIENTS IN SCOLIOSES

Andrzej Dyszkiewicz, Zygmunt Wróbel
Institute of Computer Science, Silesia University
Department of Biomedical Computer Systems, ul. Będzińska 39, 41 – 200 Sosnowiec, Poland
Andrzej1074843@pharmanet.com.pl, wrobel@.us.edu.pl

Józef Opara
Rehabilitation Department University of Physical Education, Katowice, Poland
jozefopara@interpc.pl

Keywords: Scoliosis asymmetry coefficients, lung asymmetry factor, breath asymmetry factor, pelvic asymmetry factor foot asymmetry factor.

Abstract: The work presents a clinical outline of stature defects and scoliosis as well as the contemporary methodology behind the thorax, spine and leg bone radiogram measurements. In order to increase the repeatability of the results and to create computer records which support monitoring records of scoliosis, an algorithm for the process of radiological image was developed. It automatises the time consuming process of measuring and processing data by the doctor. The image processing is initiated by an interactive procedure where key points of biological structures are marked with a cursor. Other measurements are done automatically. The algorithm is also an attempt to use the author's modification for measuring the geometry of the spine and thorax, which increases precision when compared to the methods by Cobb, Fergusson and Gruca. Results of radioplan-metric investigations compared with a system for analysing the trajectory of respiratory motion and the asymmetry weight distribution system in the foot have been presented. A mathematical analysis of thorax and bone radiogram geometry combined with the results of thorax trajectory movement enable the creation of individual patient symmetry indices with a description of the monitoring process of the disease.

1 INTRODUCTION

Side curvatures of the spine are a huge individual and social problem. The frequency in which this problem occurs is different in various populations and alternates between 3-15% of children and young people. During the last several years there has been a clearly increasing tendency in the number of people suffering from this disease. Scoliosis is a severe impediment of posture, which is accompanied by secondary alterations in the blood circulation and in the respiratory system. The alteration of the aforementioned systems leads to the limitation of the general efficiency of the patient. Finally it can cause early disabilities and can also shorten one's life.

Be advised that papers in a technically unsuitable form will be returned for retyping. After returned the manuscript must be appropriately modified. The three-dimensional body system in the standing position, which functions as a biomechanism with a wide range of freedom, protects against gravitation in the conditions of unstable equilibrium, and it should be considered in a dynamic sense. As a starting point it is necessary to consider the degree in which the body's physiology in terms of bones, joints and ligaments has been fulfilled. The correct build of the skeleton, passive stabilisation and proper movement in the range of individual elements, are essential. These elements create merely potential possibilities of taking on and maintaining the correct posture, however its image depends on the function and efficiency of the central nervous system. Currently it is claimed that assuming and sustaining the correct position is the same motion task as any other movement activity.

Producing and establishing a proper unconditioned co-ordination reflex requires even millions of conditions to be fulfilled; however the quality of created habits depends on conscious action and inborn predispositions. Gradually worked out, more simple movement abilities create a base for the following predispositions, which are built-in into more complex movement patterns. An established system of individual segments of the body, as well as balance do not form a static condition, but one that oscillates around the balance point, with a tendency to optimisation. Secondary disturbance of posture control is caused by a slightly different mechanism. It is probable that as a result of disturbance of brain functions, deviation of posture from the pattern recognised as a correct can occur. It should be noticed that while comparing the results of the application of advanced technology with a system of a few relatively connected simple methods supplemented by computer techniques, a significantly higher acuteness and peculiarity, than in the case of expensive novelties is achieved. Taking into account the unfavourable financial realities of Polish science there is a growing interest concerning the low-outlay adaptations of more simple methods. A system of mathematical correlation between different methods, which describe a chosen phenomenon, uncovers a new multidimensional space of description, characterised by a higher level of specificity.

One of the methods used in many centres as a parameter, which monitors the development of scoliosis, is the spirometer experiment. Unfortunately, from the bioengineering point of view it can be noticed that the two lung spaces coupled by an angle of scoliosis are the source of an averaged volumetric parameter and can be described as a compensatory structure. The increase in the value of the angle of scoliosis leads to a decrease of volumetric space on the chord site and a proportional increase of volume on the curve site. Proportions are preserved best with a slight scoliosis, which does not exceed $15\text{-}18^0$. In such a situation equilibrium is established. The pressure of the lung is compensated by the emphysema of the second lung. During the increase in the angle of scoliosis the summative volumetric parameter decreases. Therefore, cases where a spirometric parameter represents an angle of scoliosis the case is usually so advanced that it should be treated surgically. The non-invasive character of simple postural-metric methods and precision of planimetric radiography are connected by a method, which consists of a multi-segmental system for evaluating the trajectory of thorax movement. A system of tapes and transducers braiding the thorax enables the estimation of movement in all of its areas. Based on the knowledge of motion biomechanics' index values which are dependant on the scoliosis angle, rib movement and activity of respiration muscles, a quantified body symmetry image can be obtained. Clinical cases linked with the occurrence of certain movement asymmetries can be induced through various illness mechanisms such as limb shortening, shoulder syndrome and the like. The application of an integrated system, which evaluates the basic parameters of geometry of the thorax by the planimetric method and repeated bio-engineering (Dyszkiewicz at all 1999) and spirometry provide effective and credible supervision in every stage of the disease. The image is made at the beginning of research. An analysis of the images of a patient's body carried out by means of a neural net implemented into the computer provides geometry (mainly symmetry) indicators of selected parameters. Based on these parameters the parameters of a three-dimensional trajectory of the respiratory system are compared. Monitoring the capacity of expiration in scoliosis is of a particular diagnostic importance when the parameter values fail to increase proportionally to age. This happens when a limiting value of scoliosis is achieved which is qualified to surgical treatment.

2 THE AIM OF THE STUDY

The aim of this compilation is to find a practical application of contemporary measurement methods of the side curvature of the spine to construct a practical algorithm and easy to use multipart software. The following questions were posed:

1. Does using methods (LAF, PAF, FAF, CA, FA, GA, LCC) make it possible to differentiate between the parameters of healthy and scoliotic people ?
2. Are the new, planimetric coefficients LAF, PAF, FAF (in scoliosis) well correlated with the traditional, measurement systems CA, FA, GA (Cobb, Fergusson, Gruca) ?
3. Are the planimetric coefficients LAF, PAF, FAF, CA, FA, GA (in scoliosis) better correlated with the traditional, spirometry test LCC or breath asymmetry analysing system (produces breath asymmetry factor BAF) ?

2.1 Group of Patients

Patients suffering from thorax and spine trauma, hypertension, collagen and asthmatic disease, diabetes, taking vascular medication, having frostbites and after injury to upper extremity were excluded from the study. The examinations were carried out in the following group of patients: (1) Examined group (A) - consisted of 16 women, average age 32,9±4,6 years and 9 men aged 34,7±6,3 years, with right-thorax scoliosis. (2) Control group (B) - consisted of patients with normal spine (treated in hospital for gastric illness), 15 women and 9 men, average age 35,7±5,8 years

2.2 Methods

In the first part the measurement algorithm conducts geometrical measurements according to Cobb's and Ferguson's recommendation. In the second part the author's own modification of the process is used. It is based on measuring the torsion and angle of scoliosis on the level of every vertebra with results projected on two vertical charts placed on both sides of the radiogram (Dyszkiewicz at all. 2001) The proposed method enables a partial observation of scoliosis on the background of averaging parameters. The analyser of the radiograms works with a spirometer and a device to evaluate the trajectory of the respiratory system in the thorax. The obtained multi-parameter of patients after long-term observations significantly helps to achieve a more accurate evaluation of the progression or regression of a disease (fig. 2, 3).

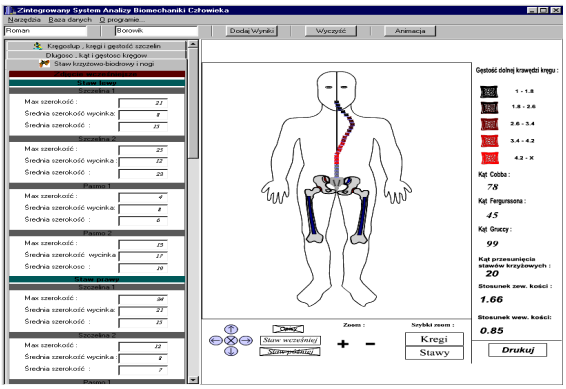


Figure 1: Expert program automatically detecting bone modification in scoliosis: (1) vertebrae torsion angle modo Gruca, Cobb, Fergusson, (2) cuneiform vertebrae deformations, (3) geometry and bone density distribution, (4) hip geometry, (5) sacro-iliac joint geometry.

The researchers used a prototypical diagnostic device, consisting of 4 elastic tapes embracing the chest, connected with converters of the path and an analogue-digital converter enabling the transmission of data through a parallel port to the "respiratory path" software which made it possible to monitor the oscillatory motion of the right and left lungs.

2.3 Results

Results of investigations (asymmetry coefficients CA, FA, GA, LAF, BAF, PAF, FAF, LCC) included in tab. 1. Patients described by asymmetry coefficients CA, FA, GA, LAF, BAF, PAF, FAF, LCC) show completely different values in group of sick patients (A) and in control group (B). While analysing table 1 we can clearly notice that in scoliosis the level of asymmetry of newly inserted coefficients LAF, BAF, PAF and FAF is comparative with coefficients based on Cobb's, Fergusson's, and Gruca's methods and clearly higher than the coefficient based on the LCC breath volume of lungs. Moreover, it can be observed that LCC in group (A) is much different from the value in the group of healthy people (B).

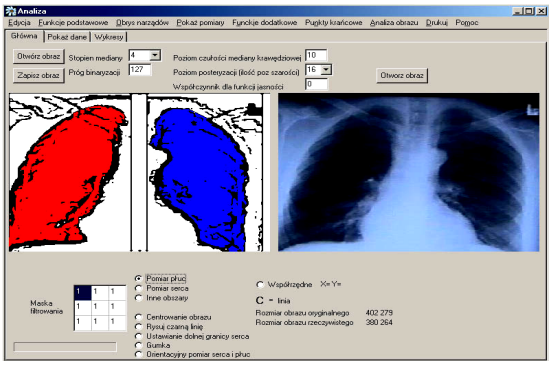


Figure 2: Asymmetry area of lungs measurement.

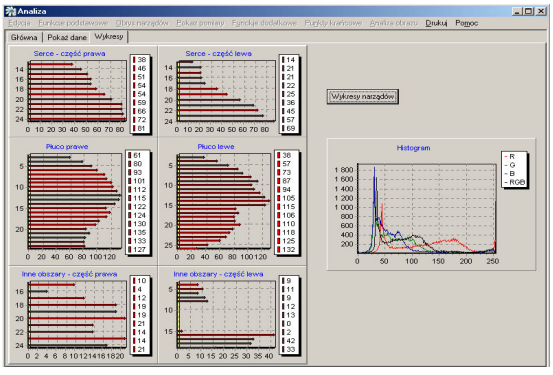


Figure 3: Graphs of ungs area.

Table 1: Asymmetry coefficients.

	Cobb angle CA	Fergusson angle FA	Gruca Angle GA	lung asymmetry factor LAF	breath asymmetry factor BAF	pelvic asymmetry factor PAF	foot asymmety factor FAF	lungs capacity coefficient LCC
Group A (n=25)	0,684 ± 0,175	0,616 ± 0,168	0,651 ± 0,181	0,752 ± 0,121	0,711 ± 0,176	0,764 ± 0,193	0,694 ± 0,185	0,876 ± 0,214
Grop B (n=24)	0,973 ± 0,181	0,955 ± 0, 114	0,949 ± 0,136	0,813 ± 0,164	0,922 ± 0,178	0,875 ± 0,187	0,935 ± 0,135	0,821 ± 0,189
CA (A - B)		P < 0,01	GA (A - B)		P < o,o1	BAF (A - B)		P < 0,01
FA (A - B)		P < 0,01	LAF (A - B)		P < 0,05	PAF (A – B)		P < 0,05
FAF (A – B)		P < 0,01	LCC (A – B)		P < 0,5			

Table 2: Correlation table of Fergusson angle with Cobb angle, Gruca angle and FAF, BAF, PAF, LAF, LCC in decreasing relation.

	FA/CA	FA/GA	FA/FAF	FA/BAF	FA/PAF	FA/LAF	FA/LCC
Gr. A (n=20)	0,93	0,915	0,89	0,81	0,76	0,72	0,41

3 CONCLUSIONS

1. The methods (LAF, PAF, FAF, CA, FA, GA, LCC) used in this investigation make it possible to clearly differentiate between the parameters of healthy and scoliotic people
2. New, planimetric coefficients LAF., PAF, FAF (in scoliosis) have good correlations with traditional measurement systems CA, FA, GA (Cobb, Fergusson, Gruca)(tab. 2).
3. The planimetric coefficients LAF, PAF, FAF, CA, FA, GA (in scoliosis) have better correlations with breath asymmetry analysing factor BAF in comparison with traditional, spirometry test LCC (tab. 2).

4 DISCUSSION

The contemporary diagnosis and monitoring of the evolution of scoliosis in cheap screening evaluation is based on a physical test assisted with plumb-line, measure tape and a goniometer. Tests of averaged spirometric parameters, which have been administered for many years, have pointed to a phenomenon where the loss of capacity of one lung, which is the result of hypopnoe, is counteracted by hyperpnoea of the other lung often leading to tests producing normal values. It is often only after the angle of curvature passes the 30’ mark that evident pathology is registered. Displaying a child on an x-ray, with the aim to determine a single angle of curvature seems very controversial. The main aim of this work was to expand the range of methods used for defining patients with scoliosis by adding to the already known methods CA, FA and GA – the new planimetric coefficients of the chest LAF, pelvis PAF, load decay of the foot test FAF and breathing track BAF. These tests were carried out on a small group of patients and have to be treated tentatively; nevertheless it is possible to notice a significant difference in the results between sick group (A) and control group (B). The correlation between CA, FA, GA / LAF, PAF and CA, FA, GA / FAF, BAF was also very good and clearly higher from the correlation with the traditional spirometric test which produced a result of only 0.41.The study results distinctly recommend inserting some extra LAF and PAF determinants to estimate standard radiological photos and to expand their range by a simple FAF podoscopic test and breath asymmetry analyses system estimating BAF. It can be clearly seen that an assessment of the breath asymmetry truck of the chest reveals more sensitivity than a traditional evaluation of the breath-volume.

REFERENCES

Dyszkiewicz A, Sapota G, Imielski K. Perspectivy automatizace plynule menitelne zateze pro krcni trakci. Rehab Fizik Lek 1999,4(6), 119

Dyszkiewicz A, Kuna J, Koprowski R, Wróbel Z. Simplified analysis of spine thermovision picture in diagnostics of scoliosis. Acta Bioeng Biomech 2001, 1,93.

DESCRIBING CRYPTOBIOSIS AS A TIME BASED PROTECTION SYSTEM USING PETRI NETS

Bengt Carlsson [a], K. Ingemar Jönsson [b] and Keith Clark [c]
[a] *School of Engineering Blekinge Institute of Technology Soft Center S-37225 Ronneby, Sweden bca@bth.se*
[b] *Department of Mathematics and Science, Kristianstad University,S-291 88 Kristianstad, Sweden. ingemar.jonsson@hkr.se*
[c] *Department of Computing, Imperial College, London klc@doc.ic.ac.uk*

Keywords: Cryptobiosis, Petri net, Time based security model.

Abstract: Cryptobiosis represents the state of a living organism when it shows no visible signs of metabolic life, but maintains a capacity to return to an active, metabolic state. This peculiar state, although known from a wide variety of organisms, has received little attention from a theoretically biological perspective. A description based on a Petri net setting and a time based security model is proposed. In order to protect against a prolonged exposure time, the pathways for chemical reactions involved must fulfil their actions during a limited detection and response time to fulfil the protected state of entering/leaving cryptobiosis.

1 INTRODUCTION

Within biology the term cryptobiosis, or hidden life, represents the state of an organism when it shows no metabolic signs of life, while still maintaining a capacity to return to normal metabolic activity (Keilin 1959, Clegg 2001). Organisms with a capacity to enter cryptobiosis are found in a variety of prokaryote, plant, and animal taxa, and has evolved as an adaptation to survive under temporally hostile conditions, e.g., in very dry or cold environments (Alpert 2005). One of the most common environmental agents inducing cryptobiosis is desiccation, leading to a variant of cryptobiosis called anhydrobiosis. The cryptobiotic state is characterised by temporally arrested metabolism, growth, reproduction and senescence (Keilin 1959, Crowe 1971). The lack of metabolism during cryptobiosis can be described as temporary "death" representing a unique biological state between life and death, a potentially reversible death (Neuman 2006).

The biochemical and physiological mechanisms allowing cryptobiotic organisms to survive in an ametabolic and (in the case of anhydrobiosis) more or less completely dry state remain poorly understood (Schill et al. 2004, Watanabe et. al. 2002, Guppy 2004). Also, the problems connected with extreme desiccation and a complete shut-down of the metabolic machinery have rarely been analysed theoretically. However, some criteria are necessary to fulfil for successful cryptobiosis to take place. For instance, the organism must either prevent cellular damage at the entrance of cryptobiosis and during the cryptobiotic state, or be able to repair the damage that is potentially expressed when reactivated. It must also be able to respond to one or more triggering signals connected with the cryptobiotic period. We will here describe a possible approach to a protection system during an induction phase, a dormancy phase and a reactivating phase of cryptobiosis.

In order to return to an active life, an organism in cryptobiosis needs to interpret signals from the environment including reactivating itself from the ametabolic state, i.e. recovering original biological functions despite the lack of basic metabolic machinery. The organism must be "raised from the dead", by responding to an environmental signal announcing favourable life conditions, e.g. a drop of water for a dehydrated organism. It has been suggested that cryptobiosis entails reversible computation with a bootstrapping involving a recursive hierarchy (Neuman 2006). This is similar to models within computer science including creating more complex tools from simple tools (bootstrapping), sub and superclasses within object oriented programming (hierarchies), and logic programming using recursive functions. We will here use a Petri net as a tool for describing the reversible processes of cryptobiosis, without going into the details of the metabolic pathways involved.

Biological systems modelled by Petri nets were introduced by Reddy et al (1993). By using

Petri net it was possible to dynamically model the intrinsic behaviour of e.g. metabolic pathways within a cell. In Heiner et al (2004) the metabolic pathways conducting an apoptosis (genetically programmed cell death) is modelled using a qualitative Petri net.

In section 2 the different phases of cryptobiosis are outlined followed by a description of a time based protection system. Next, a Petri net is introduced for modelling metabolic protection and finally a discussion and concluding part sum up the proposed description.

2 CRYPTOBIOSIS

Cryptobiosis may be induced by several environmental factors e.g. low temperature (cryobiosis), lack of oxygen (anoxybiosis) or lack of water (anhydrobiosis). All of these factors may force the organism to an arrested metabolism. Here, we will use anhydrobiosis as an example to describe cryptobiosis within a Petri net setting using a time based security model.

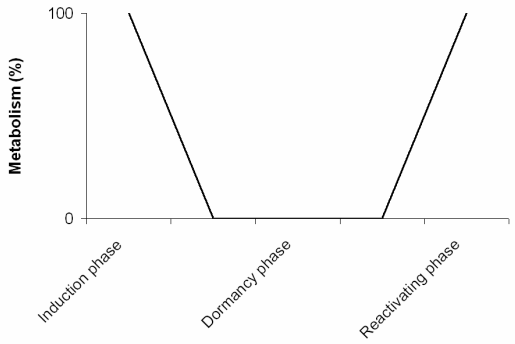


Figure 1: An idealized figure of changes in metabolism during the induction, dormancy, and reactivating phases of cryptobiosis. A more realistic curve would likely involve non-linear patterns of metabolic changes during induction and reactivation.

Figure 1 shows the three general phases connected with cryptobiosis, where the induction phase prepares the organism for an ametabolic state as a result of e.g. desiccation. Note that the factor inducing changes in metabolic rate, e.g. reduced hydration level, is not shown in Figure 1. During the dormancy phase the organism is inactive due to a lack of metabolic activity, but continued desiccation may occur because metabolism will stop well before the animal is completely dehydrated. The reactivating phase involves both the reverse processes of the induction phase and a repair process depending on damages arisen during the induction and dormant phases.

The pattern of water loss at the induction of cryptobiosis as a result of dehydration (anhydrobiosis) has been well documented (Wright 1989, Wharton 1996). Initially the evaporation of water is high, but at some point the organism has mobilised a first protection mechanism (mainly based on morphological changes) which dramatically reduces the rate of further dehydration. Wright (1989) termed this point the "permeability slump". After this point, the rate of dehydration is much reduced. The remaining time, until the organism has lost so much of its water that metabolism is arrested, is decisive of whether the organism will enter a cryptobiotic state or die. During this time, the organism must mobilise the mechanisms that should protect it as it approaches the dry anhydrobiotic state. Figure 2 describes the temporal change in hydration level of an anhydrobiotic organism exposed to a desiccating agent.

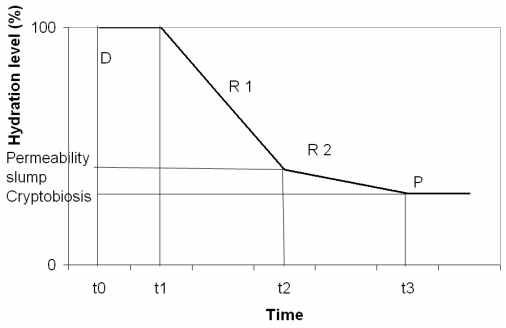


Figure 2: Patterns of water loss over time in an anhydrobiotic organism exposed to a desiccating agent, with phases of detection (D) and responses (R_i), creating a state of cryptobiotic protection (P).

In both computer science and in cryptobiotic organisms, a detection mechanism must recognize the attack or stress input, after which different response mechanisms must be activated (R1 and R2 in Figure 2). In cryptobiotic organisms these preventing mechanisms together form the protection system which must be activated before the metabolism disappears.

3 A TIME BASED PROTECTION SYSTEM

To describe, in general terms, a cryptobiotic protection system, a time based security model (Schwartau 1999) may be used, where a protection

mechanism is successful only as long as the sum of the time needed by the detection mechanism and the response mechanism does not exceed the time limit of the protection.

In this model the amount of exposure time E, which may be positive or negative, determines the outcome of an external attack. The exposure time defines a time period under which the organism is exposed to the environmental stress without a mobilized protection system. The exposure time is determined by three factors; a protection system P, a detection mechanism D, and a response mechanism R, see Figure 2. The purpose of a protection system, independent of being within computer science or life science, is to protect involved assets. This is valid for a certain amount of time, i.e. the involved assets lose their values or an organism loses the opportunity to enter a cryptobiotic state. If detection and response time is short enough, depending on the current attack strength, the protection system will not be invaded or exposed. So a negative exposure time means that the protection mechanism manages to keep the system secured during the time when detection and responses are established.

Let ΔP denote the duration of protection of a system caused by an attack with strength a starting at time t_0 and with protection strength s. Let ΔD denote the time it takes to detect an attack with strength a under given circumstances. Finally, let ΔR denote the time it takes to implement sufficient measures to eliminate the negative consequences of the attack with strength a after the signal has been detected. Then

$$\Delta E(a,s,t_0,t) = \Delta D(a,s,t_0,t) + \Delta R(a,s,t_0,t) - \Delta P(a,s,t_0,t) \qquad (1)$$

Equation (1) captures exposure time $\Delta E(a,s,t_0,t)$ as a relation between the three mechanisms above and the time intervals t_0 and t where $t>t_0$. The system will be safe if and only if $\Delta E<0$, i.e., if the system is not exposed to an attack without a protection system.

As already pointed out in Figure 2, a protection system consisting of detection, responses and protection are outlined for the induction phase. During the first period of time (t_0 to t_1) the organism has not yet started to lose its internal water, but a detection mechanism (D) recognizes the presence of a stress signal. This may be, e.g., the evaporation of surrounding water. In the next interval (t_1 to t_2), the organism starts to lose its body water and a first response mechanism (R_1) is mobilised. When the establishment of this mechanism is completed (at t_2), the rate of dehydration is much reduced, but continues until the organism has reached a level of hydration at which metabolism stops. Since metabolism is arrested well before the organism is completely dry (Clegg 1986), the loss of water may continue also after the cryptobiotic state has been reached. In the interval between t_2 and t_3 the second response (R_2), representing the biochemical preparations for the dry cryptobiotic state, takes place. If these preparations are successful, i.e., if the second protection system has been established properly, the organism enters cryptobiosis. The time available for the necessary actions (detection + responses) will be determined by the rate of desiccation from t_0 to t_3, influenced in part by the strength of the desiccation agent, and in part by the ability of the organism to reduce the effect of this agent. If detection or response is delayed the protection system will not be ready, and the organism will die or be damaged before reaching a state of cryptobiosis. Equation (2) formalizes this organism system:

$$\Delta E(a,s,t_0,t_3) = \Delta D(a,s,t_0,t_1) + \Delta R_1(a,s,t_1,t_2) + \Delta R_2(a,s,t_2,t_3) - \Delta P(a,s,t_0,t_3) \qquad (2)$$

The organism will enter a state of cryptobiosis if ΔE is negative, i.e. if the time needed to establish the protection mechanism ΔP exceeds the sum of the time for the detection ΔD and response mechanism ΔR_i. The main difference in the interpretation of Eq. (1) (representing a computer science model) and Eq. (2) (representing a biological model) is that in the former model the protection system is present already at the beginning, with an expected time of function. In the biological model, the protection system is not described as present from the start, but is established only after the stress agent is detected, with an expected time of establishment (ΔP). However, from the perspective of P, the responses R_1 leads to an intermediate protection system (activated at the "permeability slump") that conceptually resembles that of computer systems, since its location in Figure 2 defines a remaining time within which the R_2 responses must take place.

Processes during the reactivating part must include both activating the reversible processes of the induction phase, and repairing of damage caused by positive exposure time during induction and dormancy phases.

4 USING A PETRI NET FOR MODELLING METABOLIC PROTECTION

A Petri net consists of nodes and arcs where nodes are of two types; places p_i and transitions t_i. Tokens

represent activity of the sequence of transitions. Places, represented by circles, are passive system elements while transitions represented by boxes are active system elements, e.g. in a metabolic setting chemical compounds and chemical reactions. Arcs connect nodes of different type. The activity of a Petri net is a flow of tokens over the pathways of the net

To start with, the organism must detect the state of stress, or cryptobiosis stimuli, which in the case of anhydrobiosis is represented by the initiation of desiccation (loss of water), and the intensity of the stress factor. If the humidity surrounding the organism is very low the process towards a dry and cryptobiotic state will be more rapid, and the organism then has less time to prepare itself before metabolism ceases. Preparations for the dry state therefore often involve morphological changes that reduce the rate of water loss (e.g., Wright 2001). Previous research has shown that this is a critical phase, where the cell must be able to mobilize the necessary metabolic changes that allow it to maintain structural integrity and enter a dry state without lethal damage. These changes, or set of places p_i and transitions t_j in a Petri net setting, are represented as different pathways in Figure 3.

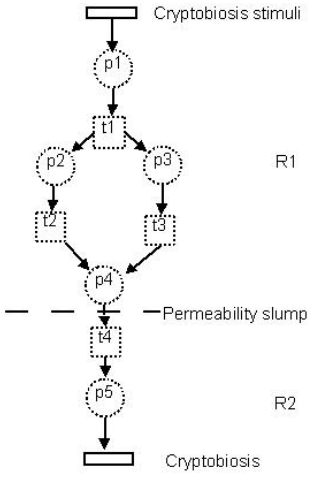


Figure 3: Induction phases of cryptobiosis. The dotted circles and squares indicate examples of metabolic pathways.

The dynamics of a Petri net allows transitions to split between multiple places or places requires more than one transition to enter a state. The number of places and transitions are not known nor are the structures of the pathways. When two arcs arrive at a place it is a synchronisation step - only when the tokens from the converging pathways have both arrived at the place, does the transition on the outgoing arc fire.

Examples of pathways are molecules (e.g., sugars, Crowe 2002) that replace structural water in cell membranes, and molecular chaperones (stress proteins) that prevent aggregation of proteins (Goyal et al. 2005). All the details involved are outside the scope of this work, but several pathways that slow down or disappear during transforming from R1 to R2 are involved.

Thus, the protection system allowing cryptobiotic survival includes several protection parameters, the induction of which relies on metabolic activity. Note that this has to be done before entering the cryptobiotic state, i.e. the protection system must act proactively.

Compared to the induction phase and the reactivating phase, which normally amounts to minutes or hours, the dormancy phase may be much more extended in time and may last for weeks, months or years (Guidetti and Jönsson 2002, Watanabe 2006).

The reactivation from the anhydrobiotic state also provides a challenge for the organism, the physiological details of which has yet to be described. At some point in time the metabolism must restart as a reaction to changed conditions, e.g., to rehydration. In principle this is the reverse function of the induction phase restarting the protection system. All the necessary biochemical components for starting up the system are present, and some of the protection components may be used for energy production or for repairing DNA and other parts of the cell.

In principle the protection system needs to be restored to its pre-conditions with one major addition, damaged parts or processes must be repaired. This is part of the existing pathways or added as new pathways enlarging the Petri net. The protection system will not be restored without successful repair facilities and, in order to have a functioning repair mechanism, some restored protection mechanisms must be present. Meanwhile there should be a replacement process where water is tied to the organism again. Finally, when the metabolism is restored to 100 percent, we are back to a fully active organism.

5 DISCUSSION AND CONCLUSION

As pointed out by Neuman (2004), cryptobiosis involves reversible computation with a suggested bootstrapping involving a recursive hierarchy. As a model for analysing the behaviour of cryptobiotic

systems, we propose using a Petri net, initiated by one or more input signals, computing a number of pathways for the induction and reactivating phases.

In computer science bootstrapping refers to a process where an initial system activates a more constant system maintaining fundamental skill. The initial process may be simple, activating a complex system step by step. Activating metabolism in a cryptobiotic organism may cause a chain reaction that step by step brings back the organism to its initial conditions. To model cryptobiosis both a Petri net setting and a time based security model is used, i.e. the chemical reactions involved and time for accomplishing the involved tasks.

The input stress to which cryptobiotic organisms are exposed could be seen as an "attack" against a system. The organism must first detect the attack and then respond to it in order to protect the system. Organisms with a fast enough detection and response will fulfil the protection mechanism's conditions. In a dynamic environment, where varying environmental conditions are presupposed, the robustness of the protection system may be modelled. This robustness is dependent of the strength of stress factor, possible threshold functions connected Petri net pathways, detection time and the success of all responses during a limited amount of time.

Currently the biological phenomenon of cryptobiosis lacks comprehensive models for describing involved processes, both at a general level and within more specific system. Such models may be found within computer science. Petri net settings may describe the different biological aspects and processes connected with cryptobiosis, and allow simulation of them. By putting more realism into the models, a future progress of both models and simulation tools may result in a better understanding of resource control within protection systems, an important issue both within life science as within computer science.

REFERENCES

Alpert, P., 2005, The limits and frontiers of desiccation-tolerant life. *Integr. Comp. Biol*. 45: 685-695.

Clegg, J.S., 1986, The physical properties and metabolic status of Artemia cysts at low water contents: the " Water Replacement Hypothesis". In: LEOPOLD, A. C. (ed.) *Membranes, Metabolism, and Dry Organisms*, pp. 169-187. Comstock Publ. Ass., London.

Clegg, J.S., 2001, Cryptobiosis – a peculiar state of biological organization, *Comp. Biochem. and Physiol. Part.B. vol 128 (4)*.

Crowe, J.H., 1971, Anhydrobiosis: an unsolved problem. *Am. Nat.* 105: 563-573.

Crowe L.M., 2002, Lessons from nature: the role of sugars in anhydrobiosis. *Comp. Biochem. Physiol*. A 131: 505-513.

Goyal, K., Walton, L.J., Browne, J.A., Burnell, A.M., and Tunnacliffe, A., 2005, Molecular anhydrobiology: identifying molecules implicated in invertebrate anhydrobiosis. *Integr. Comp. Biol*. 45: 702-709.

Guidetti, R. and Jönsson, K.I.,, 2002, Long-term anhydrobiotic survival in semi-terrestrial micrometazoans. *J. Zool.* 257: 181-187.

Guppy, M., 2004, The biochemistry of metabolic depression: a history of perceptions, *Comp. Biochem. and Physiol. Part.B. vol 139 (3).*

Heiner, M., Koch, I. and Will J., 2004, Model validation of biological pathways using Petri nets – demonstrated for apoptosis. *Biosystems 75 p. 15-28*.

Keilin, D., 1959, The problem of anabiosis or latent life: history and current concept*. Proc. R. Soc. Lond. B. 150: 149-191*.

Neuman; Y., 2004, Meaning making in the immune system, *Perspect. Biol. Med. 47*.

Neuman, Y., 2006, Cryptobiosis: A new theoretical perspective, *Progress in Biophysics and Molecular Biology 92*.

Reddy,V.N., Mavrovouniotis, M.L., Liebman, M.N., 1993, Petri net representation in metabolic pathways., *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology, AAAI Press p. 328-336*.

Schill, R.O. Steinbruck, G.H.B., and Köhler, H-R, Stress 2004, gene (hsp70) sequences and quantitative expression in Milensium tardigradum (Tardigrade) during active and cryptobiotic stages, *Journal of Experimental Biology 207*.

Schartau, W., 1999, *Time based security. Practical and Provable Methods to Protect Enterprise and Infrastructure, Network and Nation*, Interpact Press.

Watanabe, M., 2006, Anhydrobiosis in invertebrates. *Appl. Entomol. Zool.* 41: 15-31.

Watanabe, M., Kikawada, T., Minagawa, N., Yokuhiro, F., and Okuda, T., 2002 Mechanism allowing an insect to survive complete dehydration and extreme temperatures, *J. Exp. Biol. 205*: 2799-2802

Wharton, D., 1996, Water loss and morphological changes during desiccation of the anhydrobiotic nematode Ditylenchus dipsaci. *J Exp Biol.* 199: 1085-93.

Wright, J.C., 1989, Desiccation tolerance and and waterretentive mechanisms in tardigrades. *J. exp. Biol.* 142: 267-292.

Wright, J.C., 2001, Cryptobiosis 300 years on from van Leuwenhoek: what have we learned about tardigrades? *Zool. Anz.* 240: 563-582.

WEIGHTS CONVERGENCE AND SPIKES CORRELATION IN AN ADAPTIVE NEURAL NETWORK IMPLEMENTED ON VLSI

A. Daouzli, S. Saïghi, L. Buhry, Y. Bornat and S. Renaud
IMS-Bordeaux Labs, University of Bordeaux 1
351 cours de la Libération, F-33405 Talence Cedex, France
adel.daouzli@ims-bordeaux.fr

Keywords: Neuromorphic engineering, STDP, Hodgkin-Huxley model, analog VLSI.

Abstract: This paper presents simulations of a conductance-based neural network implemented on a mixed hardware-software simulation system. Synaptic connections follow a bio-realistic STDP rule. Neurons receive correlated input noise patterns, resulting in a weights convergence in a confined range of conductance values. The correlation of the output spike trains depends on the correlation degree of the input patterns.

1 INTRODUCTION

The first neurophysiology experiments on synaptic plasticity were largely inspired by Hebb's postulate (Hebb, 1949). Today, this postulate is often rephrased in the sense that modifications in the synaptic transmission efficacy are driven by correlations in the firing activity of pre- and postsynaptic neurons. Spike-timing-dependent plasticity (STDP) describes the adaptation temporal mechanisms (depression, potentiation, saturation, ...) at the level of individual spikes (Markram et al., 1997; Bi and Poo, 1998; Abbott and Nelson, 2000; Feldman, 2000; Roberts and Bell, 2002; Kepecs et al., 2002). Synapses with that kind of plasticity were found in the cortex (Markram et al., 1997), in hippocampus cells (Magee and Johnston, 1997) and in cultured cells (Bi and Poo, 1998). First studies showed the existence of Long Term Potentiation (LTP) and Long Term Depression (LTD) dependence, as functions of the synaptic weights to the time difference between the pre- and postsynaptic spikes.

More complex models were developed considering phenomenons such as previous spikes effect for the same neuron (Froemke and Dan, 2002), or the effect of synapse location (Rumsey and Abbott, 2003; Froemke et al., 2005). STDP models can also have different rules depending on the synaptic strength (Bi and Poo, 1998; van Rossum et al., 2000). These models are inspired by biophysical features. Concerning functional aspects, STDP is known to enhance the connections strength for synchronized neurons (van Rossum and Turrigiano, 2001; Song and Abbott, 2001) and is supposed to play a role in neural assembly synchronization (Singer and Gray, 1995). Depending on the shape of the STDP model, the network behavior can change, as the ratio between LTP and LTD influences the weights convergence (Song and Abbott, 2001).

Noise is considered as an interesting input in bio-realistic neural networks as it helps modeling the irregularity of real neuronal activity. Simulations showed also the impact of noise inputs on the synaptic strength evolution when driven by STDP. In (Song and Abbott, 2001), synaptic weights convergence is bimodal. STDP is applied on synapses connecting input noise spike patterns to a single spiking neuron. These input spike trains can be cross-correlated and have a Poisson distribution. With a different STDP rule, where the potentiation (LTP) depends on the synaptic strength, synaptic weights convergence is not bimodal but confined in a limited range (van Rossum et al., 2000).

Here we propose to explore, in a small neural network, the effect of correlated input noise patterns (one pattern per neuron) when a STDP rule is applied on synapses. The effect is evaluated on synaptic (between neurons) conductance distribution and on correlation in neurons' spike trains. Every neuron is attacked by an input noise pattern. These noise patterns have different levels of correlation. We use conductance-based model of cortical neurons.

STDP features are usually explored in large scale spiking neural networks, or in only one single spiking neuron. In this work, we use a 6 neurons network with a complex neuron model based on the Hodgkin and Huxley formalism (Hodgkin and Huxley, 1952). Neurons are implemented on analog VLSI circuits, and the whole simulation system is a mixed hardware-software instrumentation tool (see section 2). These

same neural chips have already been successfully used to simulate neural networks with STDP (Zou et al., 2006b). The advantage of analog VLSI for neural simulation is the speed of execution, it ensures simulations in a biological real time. Furthermore, hardware environment provides an electronic noise as in biology living cells are in a noisy environment and emulates in a way biological dispersion. In subsection 2.3, we present the STDP model we use and in subsection 2.4 the method for correlating input noise patterns. Then, in section 3, we show how hardware parameters are related to biophysiological values. In section 4, we present the simulation configuration, tools to observe the distribution of synaptic weights and the correlation of spikes, and we show results and analysis of experiments. Finally, we discuss the specifications and the results of these experiments.

2 THE SIMULATION PLATFORM

We used for the simulation a hardware implementation of a conductance-based neuron model following a Hodgkin and Huxley formalism. The implementation is done on analog VLSI circuits; the neural network connectivity is driven by a custom hardware-software system named PAX (Renaud et al., 2007). This system is embedded on a computer through a PCI interface board.

2.1 The Neurons Models

Analog VLSI circuits, model the neurons ionic currents, as described in the Hodgkin and Huxley formalism. An external capacitor connected to the circuits provides a voltage that is equivalent to the membrane potential and ionic currents channels modulate this potential. Four voltage-dependent ionic currents are implemented: $I_{Na^+}, I_{K^+}, I_{LEAK}$ and a modulating slow voltage-dependent potassium current I_M. The modeled neuron is the glutamate excitatory regular spiking neuron (Connors and Gutnick, 1990). Hardware neurons are characterized by their static parameters as time kinetics, potential offsets, conductance values (table 1), and by their functional features as f(I) curves and spike-frequency adaptation (see section 2.2). The neurons model parameter are listed in table 1. *m*, *n*, *mm* in case of activation and *h* in case of inactivation are state variables (*s*), describing the state of ionic channels, defined by: $\tau(V_{MEM})\frac{ds(t)}{dt} = s_\infty(V_{MEM}) - s(t)$ with $s_\infty(V_{MEM}) = \frac{1}{1+exp(\pm\frac{V_{MEM}-V_{OFFSET}}{V_{SLOPE}})}$.

The synapses conductance-based model is the kinetic synapse model presented in (Destexhe et al., 1994). It describes the synaptic strength as the duration of postsynaptic receptors opening (AMPA receptors for excitatory synapses). A pulse length represents the conductance increase due to the release of transmitters (Zou et al., 2006a).

Table 1: Ionic channels parameters for the implemented model, relative to a membrane area of $0.00022cm^2$.

Leak	$I_{LEAK} = g_{LEAK}(V_{MEM} - V_{EQUI})$
	$g_{LEAK} = 33nS, V_{EQUI} = -80mV$
Na	$I_{Na} = g_{Na}m^3h(V_{MEM} - V_{EQUI})$
	$g_{Na} = 11\mu S$, $V_{EQUI} = 50mV$
	$m : V_{OFFSET} = -37\,mV$, $V_{SLOPE} = 7.2\,mV$
	$h : V_{OFFSET} = -42\,mV$, $V_{SLOPE} = -4.6\,mV$
	$\tau(m) = 0.03\,ms$, $\tau(h) = \begin{cases} 3.00ms \text{ if } V_{MEM}>0 \\ 0.25ms \text{ if } V_{MEM}<0 \end{cases}$
K	$I_K = g_K n^4(V_{MEM} - V_{EQUI})$
	$g_K = 1.1\mu S, V_{EQUI} = -100mV$
	$n : V_{OFFSET} = -37mV$, $V_{SLOPE} = 11.38mV$
	$\tau(n) = 3ms$
Mod.	$I_M = g_M m(V_{MEM} - V_{EQUI})$
	$g_M = 10nS, V_{EQUI} = -100mV$
	$mm : V_{OFFSET} = -35mV$, $V_{SLOPE} = 11.4mV$
	$\tau(mm) = \begin{cases} 300ms \text{ if } V_{MEM}<0 \\ 8ms \text{ if } V_{MEM}>0 \end{cases}$

2.2 Neurons Functional Features

In the PAX system, values of stimulation currents are electronic values that can differ from one neuron to the other to trigger a same frequency. This phenomenon is due to the mismatch and variations in the VLSI circuits fabrication process. We use the f(I) curves to benchmark the circuits and tune the simulation parameters. The measured f(I) curves match the software simulations of the corresponding model. Differences exists concerning origin and scale values for the current range of the f(I) curves. Spike-frequency adaptation shape observed on raster-plots is consistent with biological data. These results are detailed in (Lewis et al., 2006).

2.3 The Neural Network Connectivity

The STDP algorithm used is based on (Badoual et al., 2006) biophysical model equation:

$$+(\omega_{ji} - \omega_{LTD})\sum_l Q[t - \tilde{t}_i(t)]\delta(t - t_{j,l})\Big] \qquad (1)$$

where ω_{ji} is the synaptic weight from neuron j (presynaptic) to i (postsynaptic). $t_{i,k}$ and $t_{j,l}$ are respectively the sets of post- and presynaptic spikes times.

P and Q are respectively the amount of LTP (potentiation) and LTD (depression) change and are given by: $P(t) = A_+ exp(-t/\tau_P)$ and $Q(t) = A_- exp(-t/\tau_Q)$. ε_k are functions taking into account spikes history of a neuron and are given by $\varepsilon_j = 1 - exp[-(t - \tilde{t}_j(t))/\tau_{\varepsilon j}]$ and $\varepsilon_i = 1 - exp[-(t - \tilde{t}_i(t))/\tau_{\varepsilon i}]$. ω_{LTP} is the maximal soft bound while ω_{LTD} is the minimal soft bound. $\tilde{t}_j(t)$ is the neuron j last spike time and $\tilde{t}_i(t)$ is the neuron i last spike time.

The STDP equation (1) is based on a precise biophysical model. Parameters for exponential constants are $A_+ = 0.1$ concerning potentiation and $A_- = 0.005$ for depression. Time constants are $\tau_P = 14.8ms$ for potentiation exponential of P and $\tau_Q = 33.8ms$ for depression exponential for Q. The eligibility ε (influence of previous spikes of a same neuron), has an exponential time constant for the presynaptic neuron $\tau_{\varepsilon j} = 28ms$ and for the postsynaptic neuron $\tau_{\varepsilon i} = 88ms$ (Froemke and Dan, 2002). This takes into account features as frequency dependence and spike triplets. The STDP algorithmic implementation and parameters are detailed in (Zou, 2006).

2.4 Correlated Input Noise Patterns

The noise inputs applied to neurons are coded as patterns generated from a Poisson distribution and correlated with a defined degree. The Poisson distribution X is obtained as follows: $X = \{x_1, ..x_n\}/x_i = N(0,1).\sqrt{m - 1/2} + m$ where N(0,1) is a normal distribution, m the average. X is converted in an absolute time pattern Y: $X = \{x_1, ..x_n\} \rightarrow Y = \{y_1, ..y_n\}$:

$$Y : y_i = \sum_{j=1}^{i} x_j \tag{2}$$

Noise input patterns are generated with one event around each Y event. The time-lap between the event Y and the pattern event is given by ε:

$$\varepsilon = N(0,1).(\alpha - 1).\frac{T}{6} \tag{3}$$

where N(0,1) is a normal distribution, T is the average period and $\alpha \in [0,1]$ is the correlation coefficient.

3 FIXING HARDWARE/BIOLOGY EQUIVALENCES

The stimulation currents applied to neurons are not directly linked to the biophysical values. The possible values are in the range 0 to 4095. We have the same inconvenience with synaptic conductances. In the hardware, the values are coded as integers that can vary from 0 to 255. To find for each VLSI neuron the correspondence with biophysical values, we developed a neuron model using the software NEURON (Hines and Carnevale, 1997) corresponding to the VLSI neuron model. The morphology of the neuron is a cylinder of 1 section, diameter 96 nm and length 73 nm. Having equivalent models, we developed a protocol to define the biophysical equivalent to the digital parameters values used in PAX.

3.1 Determining Synaptic Conductances in the Pax System

We extracted from the measurements on the PAX system a rule for converting a PAX synaptic strength value in a biophysical corresponding conductance value and conversely. We created a two neurons network (figure 1). Neuron A was stimulated by a current I_A that implies oscillations at about 8.5 Hz. Neuron B is stimulated by a current I_B that implies oscillations at about 3 Hz. Then an excitatory synapse ω_{AB} is created connecting A to B with A presynaptic to B. The synaptic strength increases from 0 to 255 for PAX, 0.02 μS for NEURON. For each weight, neurons frequencies f_A and f_B are measured (f_A doesn't change for NEURON and is near constant for PAX with standard deviation equal to 0.4 and mean value 8.7 Hz). Equations ($f(\omega) = a.\omega + b$) of the straight line fitting the experiments points are calculated for both measurements on PAX and on NEURON. The rules ($\omega_{PAX} = a.\omega_{NEURON} + b$; $\omega_{NEURON} = a'.\omega_{PAX} + b'$) give the correspondence between the biological model and the hardware parameter. Results are: $\omega_{NEURON} = (0.0943\omega_{PAX} + 3.0562)/1070.33$ and $\omega_{PAX} = (1070.33\omega_{NEURON} - 3.0499)/0.0943$.

Figure 1: Network of two neurons A and B stimulated by constant currents, respectively I_A and I_B. ω_{AB} is the synaptic strength of the excitatory synapse connecting A to B.

3.2 Determining Stimulation Currents in the Pax System

The aim here is to extract from measurements a correspondence between a PAX stimulation current value and its biophysical equivalent. The principle is to excite a neuron B by a presynaptic neuron A. A is stimulated as in previous section with a static current. The weight of synapse connecting A to B is also constant.

The B stimulation current varies in a range such that: the lowest value doesn't make B oscillating and the highest value synchronizes B to A firing. Data collected for each neuron B's stimulation current I_B are frequency of B f_B and A's frequency f_A (constant). I_A was chosen to have a frequency $f_A = 8.5Hz$. Using PAX this frequency is approximate (for 10 trials: m=8.7, SD=0.4) due to electronic noise. The simulations duration is 30 s. Frequencies are calculated between 2 s and 29 s. The equivalent synaptic weight is calculated using the rule defined in subsection 3.1. The correspondence is not exactly linear between frequency and current, but the rule that we establish provides a good approximation of the biophysical values corresponding to PAX parameters. This process has to be repeated for every neuron because of their intrinsic variability. We obtain for every PAX neuron, a correspondence rule: $I_{PAX} = a.I_{NEURON} + b$ and $I_{NEURON} = a'.I_{PAX} + b'$. For instance, the rule giving the PAX neuron number 2's biophysical current value is: $I_{NEURON} = -0.01625 I_{PAX} + 36.363$.

4 EXPERIMENTS AND RESULTS

4.1 Experiment: STDP Simulation with Correlation

The PAX system, including the VLSI neurons' was used for the experiments. The system is embedded in a computer, through a PCI interface, that computes plasticity algorithms. The computer features are: processor Intel Pentium 4®, dual core, 2.6GHz, cache: 512 Ko, SDRAM: 1Go. The operating system is the Ubuntu® Linux system. We ran a series of simulations with STDP as described in section 2.3 and noise input patterns as described in subsection 2.4. The simulated neural network comprises six excitatory neurons with all-to-all connectivity. All connections follow a STDP rule. ω_{LTP} is fixed in order to have all neurons presenting a non-bursting activity bursts. The experiment will help evaluating the STDP effects in this small excitatory network when correlated noise patterns are stimulating the neurons.

Network and Neurons Features. The neurons are stimulated by constant currents chosen from f(I) curves to maintain the membrane potential under the firing threshold. Each neuron receive additional stimulation: an input noise pattern (rate 5Hz) tuned in order to trigger an oscillating frequency lower than 5Hz (mean value 3Hz). All currents are in the biological range [0.4nA-0.5nA].

Initial synaptic weights are either null or randomized using an uniform law. Corresponding randomized conductances values, using the correspondence rules determined in section 3.1, are in the range [0nS-20nS] which corresponds to numerical values in the range [0-180]. Furthermore, a neuron receives synaptic inputs from all other neurons and projects its output to all synapses of the other neurons. The simulation lasts 360 seconds. When a neuron spikes, all the related synaptic weights are recalculated using the STDP algorithm. For data analysis, each weight change is recorded together with the timing. For each neuron, all the timing of its spikes are also recorded for further analysis.

4.2 Analysis Tools

Weight Histogram. The method used to assess weight convergence is the building of a histogram of weights distribution (see top line of figure 2). For this experiment, the encoded maximum weight value is 180. We divide the weight axis into 36 bins, thus each section corresponds to an interval of 5. The weights distribution is then calculated at the end of the simulation and normalized.

Spike Correlation Histogram. To evaluate the correlation between the neurons output firing patterns, a correlation histogram is defined (see bottom line of figure 2). The method is to divide the time axis into sections, each section corresponding to 10 ms. Spikes occurring in each time section are accumulated. As we have 6 neurons, the maximum count per section is 6 if every neuron spikes in that 10 ms window, except if a neuron spikes 2 times in the same window. We don't consider that exception here. For every possible spikes count (here from 1 to 6), the number of sections having this value is calculated. Both axis are then normalized. This provides a graph showing the spikes distribution relative to a minimal time window. If the spikes of different neurons are well correlated in that time window, distribution tend to 1, whereas distribution will be closer to 0 for uncorrelated activity.

4.3 Results Analysis

As we can see on the bin histogram of figure 2A, the weights after STDP are distributed in a limited range when the correlation of input patterns is weak ($\alpha < 0.5$, α from equation 3). When the input correlation grows, extrema values of the weights appear (e.g. with $\alpha = 0.6$ figure 2C). When the input correlation is maximum ($\alpha = 1$) then the weights distribution

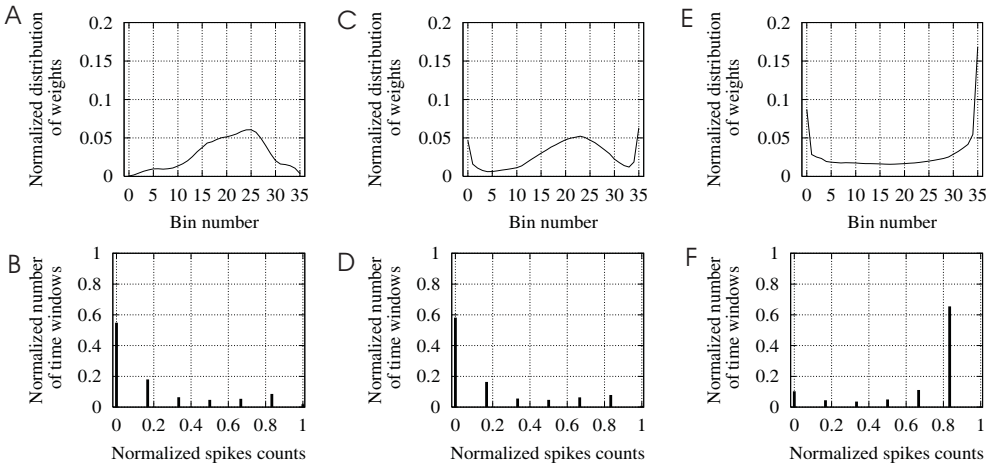


Figure 2: Simulation of input correlation's effect on the synaptic weights convergence and the spikes correlation. Time simulation lasts 360 s with initial weights randomized and a frequency of input noise around 5Hz. *Top line*: Weight histograms for a simulation with input correlation of 0.35 (A), 0.6 (C) and 1 (E). *Bottom line*: Correlation histograms for a simulation with a time window of 10 ms. The input correlations are 0.35 (B), 0.6 (D) and 1 (F).

is bimodal (figure 2E). At the same time, the correlation histogram shows that correlation of output spikes in a window time of 10 ms is weak when correlation input has $\alpha < 0.8$ (figure 2B and D), grows when α is higher. Up to a high correlation when $\alpha = 1$ (figure 2F).

In (van Rossum et al., 2000), where the STDP rule has no soft bound (ω_{LTP} and ω_{LTD}) and where LTP depends on the synaptic strength (ω_{ji}), the weights always converge in a limited range. In (Song et al., 2000), where the soft bound is introduced but with no LTP depending on ω_{ji}, the weights systematically show a bimodal convergence.

Our STDP model was simulated in (Zou and Destexhe, 2007) in single neuron configuration. In that case, all weights converge into a limited range.

The experiments we presented showed that this same STDP rule applied to a 6-excitatory neurons network lead to more complexe figures, mixing bimodal and range limited weights convergence.

5 DISCUSSION AND CONCLUSIONS

Using analog VLSI circuits for computational neuroscience is a performant solution for running simulations at biological real time. The system used can also be interfaced with real biological neurons to create a hybrid neural network (Le Masson et al., 2002). One inconvenient, as seen in section 3, is that the tuning of some parameters depends on the fabrication for parameters. The correspondence rules developed provides an estimation of biophysical values and only in a short range because of non-linearities. However, benchmarks showed us the network patterns were respected by such a simulation tool (Zou et al., 2006a).

In our experiments, we showed that the weights distribution convergence depends on correlation of input noise patterns. This convergence mixes bimodal convergence and range confinement convergence. This phenomenon is not covered by other STDP rules. We also showed that the input correlation degree influences the correlation in neurons spikes. For the next PAX system generation, STDP computation will be embedded in the hardware system. A more important number of neurons will be available with available inhibitory neurons. Experiments will be continued on more complex neural networks.

ACKNOWLEDGEMENTS

This work is supported by the European Community Grant FACETS (IST-2005-15879).

REFERENCES

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Natural Neuroscience*, 3:1178–1183.

Badoual, M., Zou, Q., Davison, A. P., Rudolph, M., Bal, T., Frégnac, Y., and Destexhe, A. (2006). Biophysical and phenomenological models of multiple spike interactions in spike-timing dependent plasticity. *Int. J. Neural Syst.*, 16(2):79–98.

Bi, G. and Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike tim-

ing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–10472.

Connors, B. and Gutnick, M. (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences*, 13:99–104.

Destexhe, A., Mainen, Z., and Sejnowski, T. J. (1994). An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Computation*, 6:10–14.

Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, 27:45–56.

Froemke, R. C. and Dan, Y. (2002). Spike-timing-dependent plasticity modification induced by natural spike trains. *Nature*, 416:433–438.

Froemke, R. C., Poo, M., and Dan, Y. (2005). Spike-timing-dependent plasticity depends on dendritic location. *Nature*, 434:221–225.

Hebb, D. O. (1949). *The Organization of Behaviour*. John Wiley & Sons.

Hines, M. L. and Carnevale, N. T. (1997). The neuron simulation environment. *Neural Computation*, 9:1179–1209.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544.

Kepecs, A., van Rossum, M. C. W., Song, S., and Tegner, J. (2002). Spike timing dependent plasticity: common themes and divergent vistas. *Biological Cybernetics*, 87:446–458.

Le Masson, G., Renaud-Le Masson, S., Debay, D., and Bal, T. (2002). Feedback inhibition controls spike transfer in hybrid thalamic circuits. *Nature*, 417:854–858.

Lewis, N., Bornat, Y., Alvado, L., Lopez, C., Daouzli, A., Levi, T., Tomas, J., Saghi, S., and Renaud, S. (2006). Pax : un outil logiciel / matériel d'investigation pour les neurosciences computationnelles. In *NeuroComp*, pages 171–174.

Magee, J. C. and Johnston, D. (1997). A synaptically controlled, associative signal for hebbian plasticity in hippocampal neurons. *Science*, 275:209–213.

Markram, H., Lubke, J., Frotscher, M., and Sackmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215.

Renaud, S., Tomas, J., Bornat, Y., Daouzli, A., and Saighi, S. (2007). Neuromimetic ICs with analog cores: an alternative for simulating spiking neural networks. In *InternationaI Symposium on Circuits And Systems*, pages 3355–3358. IEEE.

Roberts, P. D. and Bell, C. C. (2002). Spike timing dependent synaptic plasticity in biological systems. *Biological Cybernetics*, 87:392–403.

Rumsey, C. C. and Abbott, L. F. (2003). Equalization of synaptic efficacy by activity- and timing-dependent synaptic plasticity. *The Journal of Neurophysiology*, 91:2273–2280.

Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18:555–586.

Song, S. and Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32:339–350.

Song, S., Miller, K., and Abbott, L. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3:919–926.

van Rossum, M. C. W., Bi, G.-Q., and Turrigiano., G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20:8812–8821.

van Rossum, M. C. W. and Turrigiano, G. (2001). Correlation based learning from spike timing dependent plasticity. *Neurocomputing*, 38-40:409–415.

Zou, Q. (2006). *Computational models of spike timing dependent plasticity: synapses, neurons and circuits*. PhD thesis, Université Paris VI.

Zou, Q., Bornat, Y., Saïghi, S., Tomas, J., Renaud, S., and Destexhe, A. (2006a). Analog-digital simulations of full-conductance-based networks of spiking neurons with spike timing dependent plasticity. *Network: computation in neural systems*, 17:211–233.

Zou, Q., Bornat, Y., Tomas, J., Renaud, S., and Destexhe, A. (2006b). Real-time simulations of networks of hodgkin-huxley neurons using analog circuits. *Neurocomputing*, 69:1137–1140.

Zou, Q. and Destexhe, A. (2007). Kinetic models of spike-timing dependent plasticity and their functional consequences in detecting correlations. *Biol. Cybern.*, 97(1):81–97.

INSECT SENSORY SYSTEMS INSPIRED COMMUNICATIONS AND COMPUTING (II): AN ENGINEERING PERSPECTIVE

Zhanshan (Sam) Ma, Axel W. Krings and Robert E. Hiromoto
Computer Science Department, University of Idaho, Moscow, ID 83844, USA
{sam,krings,hiromoto}@cs.uidaho.edu

Keywords: Insect Sensory System, Micro Aerial Vehicle (MAV), Wireless Sensor Network, Non-cooperative Social Behaviour, Insect-Inspired Robot, Cellular Computing, Agent-based Computing, Biosensor, Dendritic Neuronal Computing, Molecular Network.

Abstract: This is the second article in a two-part series in which we briefly review state-of-the-art research in communications and computing inspired by insect sensory systems. While the previous article focuses on the biological systems, the present one briefly reviews the status of insect-inspired communications and computing from the engineering perspective. We discuss three major application areas: wireless sensor network, robot and micro aerial vehicle (MAV), and non-cooperative behaviours in social insects and their conflict resolution. Despite the enormous advances in insect vision and mechanosensory inspired robot and MAV, micro-flight emulation, motion detection and neuromorphic engineering, etc., the potential inspiration from insect sensory system is far from being fully explored. We suggest the following promising research topics: (1) A new grid computing architecture emulating the neuronal population such as the visual neurons that support the compound eyes, the PN (Projection Neurons) in AL (Antennal Lobe) or the ORN (Olfactory Receptor Neurons) from insect sensory organs (sensilla). This may be further integrated and enhanced with the dendritic neuronal computing. (2) New generation of *multimodal* wireless sensor and ad-hoc networks that emulates insect chemosensory communication. The inspiration of multimodalities in insect sensory systems also implies that there are multiple *parallel* networks operating concurrently. Furthermore, the insect chemosensory is significantly robust and dependable with built-in anti-interference mechanisms. (3) Non-cooperative behaviours in social insects may offer insights to complement swarm intelligence (inspired by cooperative behaviours) or to devise new optimization algorithms. It may also provide inspiration for proposing survival selection schemes in evolutionary computing. We suggest using evolutionary game theory to model conflict resolution in social insects, given its success in modelling conflict resolution of other animals.

1 INTRODUCTION

Organisms interact with each other and with their environments through sensory and motor systems; so do the engineered systems. Their stability and control depend on continuous sensing and actuation (Miesenbock and Kevrekidis, 2005). This argument shows the universal significance of sensory systems to both biological and engineered systems, which is particularly true to insects given insect sensory systems are one of the top four reasons contributing to their status as the most abundant organisms on earth (Ma and Krings, 2007).

Two terms often appear in bio-inspired computing: biomimetic and biomorphic. The former is more common and emphasizes the mimic or emulation of nature and the latter is more of a metaphor (Lodding, 2004). The applications we survey below largely fall into one of the two categories, but in reality, the distinction is rarely clear-cut. In some occasions, a biologically inspired approach is *recursively* applied to solve biological problems (e.g., biosensoring in section 3).

To harness the biological inspirations from insect sensory systems, being familiar with the biological aspects is necessary. We refer to the following excellent monographs (Christensen 2005, Drosopoulos and Claridge 2006), both of which are dedicated to insect sensory systems. An excellent and up-to-date monograph, rightly acclaimed by reviewers as providing "a remarkably holistic yet detailed view" on insect physiological systems including the sensory systems, should be an ideal reference for studying insect sensory system in more comprehensive context (Klowden, 2007). General knowledge on insect sensory systems can be found in an entomology textbook such as Gullan and Cranston (2005). Numerous proceedings from

symposiums and conferences on bio-inspired computing have been published since the 1990s and a significant amount of research is inspired by insects (e.g., Detrain et al. 1999, Dressler and Carreras 2007). A possible starting point, which provides an article-level review of the insect sensory systems from the perspectives of inspiring communications and computing, could be the Ma and Krings (2007). Given the extensive existing literature, which continues to accumulate faster than ever, we choose a significantly narrow scope in this article to focus on insect sensory system related topics. Even with the narrowed-down scope, it still seems impossible for us to present a comprehensive review in such a short article. Therefore, we choose to focus on three research areas and exclude the others. In addition, priority was given to the state-of-the-art review papers, monographs, and research papers representing a major category of studies (often limited to one per topic). Consequently, we have to regrettably omit a number of excellent research papers. As a minor remedy to the excluded fields, in section 3, *the other topics,* we mention five areas and a few review references about them.

2 INSECT SENSORY SYSTEMS INSPIRED COMMUNICATIONS AND COMPUTING

2.1 Wireless Sensor Networks

It seems that insect sensory systems may inspire the design of wireless sensor networking on both the node (sensor node *vs*. individual) level and network level (sensor network *vs*. insect population).

The inspiration at the individual sensor node level is the most obvious. Essentially, a robot emulation of insect vision and navigation provides a typical example for this kind of research, where each individual insect is *mapped* to an engineered sensor. Many of the neural sensory mechanisms in insects can be emulated in individual sensor design. In particular, multimodality capability is very desirable in sensor networks (Ma and Krings, 2007).

From the population perspective, potentially two types of "mappings" can be construed. The first type is the neuron population or the group of neurons behind a sensory organ such as antenna or compound eyes. This type of neuron population forms a *grid computing* infrastructure (similar to the cellular computing paradigm). The populations of ONRs (olfactory neural receptors) and PNs (projection neurons) in the olfactory system are examples of this type (Ma and Krings, 2007).

The other type of population organization *mapping* can be the population of insect individuals *vs*. *population* of wireless sensor nodes, i.e. wireless sensor network. An insect population that distributes over habitat space forms an information network. This network may depend on infochemicals (in chemosensory system) or vibrations (in audition) as "packets" communicating via air, water, or other types of substrate medium. Indeed, the infochemicals-based *wireless* communication network is probably more complex than electron-based networks. Several categories of infochemicals are involved, e.g., pheromones are utilized in intra-specific communications and allelochemicals (allomones, kairomones, and synomones) in inter-specific communications (Ma and Krings, 2007). What may be even more inspiring is that there are several *parallel* communications networks—visual, olfactory, auditory, etc.—in an insect population, or the so-called multimodality sensory. All of the sensory networks are *wireless* except for the taste sensory network. This is essentially the demonstration of multiple modalities at the population level.

In terms of sophistication and functionalities, no other organisms may match insects in the chemosensory systems. The differences between the insect chemosensory *wireless* network and the engineered wireless network lie in message encoding (infochemicals vs. radio frequencies) and computing node (insect brain vs. microchip). The research of insect sensory systems may inspire the engineering of reliable and secure wireless sensor networks. Obviously, the insect sensory *wireless* network is operated under heterogeneous and unstable natural environments. The network has to deal with possible exploitations by other species, which may be their competitors or natural enemies. For example, the natural enemies may try to find their prey by following the infochemicals, and the insects may release interference infochemicals to confuse their competitors. This is similar to malicious intrusions in computer networks.

2.2 Insect-Inspired Robots and Micro Aerial Vehicle (MAV)

The study of the aerodynamics of insect flight was conducted as early as the 1950s. Grasshoppers and flies seem to be the most common model insects. Both walking (including crawling) and flying robots based on insects have been developed. Insect sensory systems, mainly vision and mechanosensory, have offered inspiration for those

designs. It can be said that the research of insect-inspired flight has been the most intensive and extensive field studied among all insect-related engineering studies.

Micro Aerial Vehicles (MAV), also known as Mini Aerial flight Vehicles, have been studied for over a decade. An MAV is based on UAV (Unmanned Aerial Vehicle) technology, but there are significant differences. According to DARPA's definition, an MAV has a wingspan of less than 15 cm. It turned out that the 15 cm is an interesting threshold to separate two types of flights: flapping flight (*micro-flight*, used by insects) vs. fixed wing soaring flight (Pornsin-sirirak et al. 2001).

At least seven laboratories started insect-inspired robots research in approximately the same period about a decade ago. The Biomimetic Millisystem at U.C. Berkeley has been developing the so-called minimally-invasive flying robots, weighing 0.1g, using insect-inspired wing kinematics (Wood et al. 2005, Steltz 2005). The group at CalTech's (Pasadena, CA) Micromachining Lab focused on the design and manufacturing of flight wings for MAV. For example, they developed the first MEMS-based (Micro Electro Mechanic Systems) wing technology with titanium-alloy metal as wingframe and parylene-C as wing membranes (Pornsin-sirirak et al., 2001). The "Entomopter" is a multimode (flying/crawling) robot designed by the joint team of Georgia Tech Research Institute (GTRI) and Cambridge University. The effort has been made to develop an Entomopter-based Mars surveyor (Michelson, 2002). The Biorobotic Vision Laboratory at the Australia National University has focused on the insect vision-driven behaviors and their inspiration for machine vision, as well as visually guided robots (Srinivasan et al., 2001, 2003). Their researchers, in cooperation with the Jet Propulsion Laboratory at Cal-Tech and NASA, have developed robots for Mars exploration based on the study of ocelli of dragonflies. The design of Mars exploration robots has taken inspiration from the unique skills of dragonflies in navigation, hazard avoidance, altitude hold, stable-flight, terrain-following and smooth deployment of payload (Thakoor, 2003). The Center for Intelligent Mechatronics at Vanderbilt University studied Mesoscale Crawling Robots based on insect model (Lobontui et al., 1999). CAVIAR is a European Commission funded project to develop a multi-chip vision system based on Address-Event Representation (AER) communication of spike events (http://www.ini.uzh.ch/~tobi/caviar/). This system emulated biological visual pathways (Liu et al. 2002). *Fly-by-Sight-Microrobots* is a project headed by Nicolas Franceschini in France. His team developed neuromimetic robots by emulating the fly's compound eyes (www2.cnrs.fr/en/582.htm).

Besides the previous groups' comprehensive research projects, quite a few researchers have conducted relatively ad-hoc studies in the field. Motamed & Yan (2005) is a review of insect-inspired micro-flight. Ma and Krings (2007) reviewed more case studies in insect-inspired robots and MAV.

2.3 Non-Cooperative Behaviours in Social Insects — Conflicts Resolution

Non-cooperative behaviors in social insects are contrary to the cooperative ones that have inspired swarm intelligence and similar algorithms, also referred to as ants colony optimization algorithms. The reason we single out this type of insect behavior is an intuitive argument: If the solution for the opposite side of a problem is inspiring, one may be able to get the solution by conducting inverse transformation. This is often true in optimization.

The society of social insects, like any society, is never a perfect world. The dominant organization of the insect societies such as bees, ants and termites is the caste system, and individual *rights* are often not fully protected. Two major conflicts exist in social insects: (1) direct reproduction rights and (2) the manipulation of fellow colony members. Ratnieks and Foster et al (2006) reviewed five major reproductive conflicts in insect societies, including: (1) sex allocation, (2) queen rearing, (3) male rearing, (4) queen-worker caste fate, and (5) breeding conflicts among totipotent adults. These reproductive conflicts exist widely in the colonies and sometimes have dramatic effects on the colonies. Three essential mechanisms: kinship, coercion, and constraint typically jointly limit the effects of conflicts and often the reproductive conflict is resolved totally. The *inclusive fitness theory* has been proposed to explain both cooperation and conflict. Essentially some individuals of a colony relinquish their direct reproductive rights to help rear and defend the offspring of other colony members. A major factor in conflict resolution is the kinship, since the great relatedness suppresses the incentive to be selfish. Whether or not the pheromones, which play crucial roles in cooperative behaviors, are involved in conflict resolution is still unknown, and neither are the genes affecting conflict resolution (Ratnieks and Foster et al. 2006). There has been no modeling research of the conflict resolution in social insects.

Whether or not pheromones are involved in the conflict resolution is really not important for their potential inspiring in devising new computation strategies or extending the existing swarm intelligence. (The latter is based on pheromone-regulated cooperative behaviors.) We see three potentially rewarding explorations. (1) Extending swarm intelligence. In real world populations, both cooperative and non-cooperative (conflict resolution) mechanisms exist simultaneously and the successful resolution of conflict may enhance cooperation. Therefore, introducing conflict resolution into swarm intelligence algorithms should make the algorithms match biological mechanisms more consistently. Cooperative behavior is essentially a positive feedback mechanism, and non-cooperative behavior often acts as the negative feedback mechanism. A system should become more stable with both types of feedback regulations. Certainly, what we suggest here is just a conjecture. (2) The mechanisms of conflict resolution may be inspirational for designing survival selection mechanisms in evolutionary computation, or extending the existing survival selection schemes. (3) Mathematical modeling of the conflict resolution in social insects has not yet been explored. Given the dominant role of evolutionary game theory in modeling conflicts resolution in other animals (Maynard-Smith 1982), it makes great sense to apply evolutionary game theory first. Obviously, the studies of (2) and (3) should be compared to inspire each other, since the topic of (2) is essentially an evolutionary computation issue and that of (3) belongs to evolutionary biology.

3 THE OTHER TOPICS

Insect Vision Inspired Motion Detection and Neuromorphic Engineering. This topic was addressed in the first article of this two-part series (Ma and Krings 2007), since it was more convenient to discuss it in the context of insect vision sensory systems. Given its extreme importance, we include the following brief summary.

One field that has made enormous progress in recent years is the motion detection of insect eyes and their applications to bio-inspired robot sensors. This is one area of neuromorphic engineering. *Parallel* and *analog* are two trademark properties of insect neural systems. It is now possible to design and manufacture a fully integrated neuromorphic olfaction chip (Liu et al. 2002, Stocker 2006, Koickal et al. 2007). A possible reason for the advancement is that motion-detection neurons are some of the largest in insect vision systems and easy to observe (Rind, 2005). Rind (2005) summarized three types of contributions where man-made vision systems are based on insect vision system: (1) Bio-inspired circuits embedded in the control structure of mobile robots. Examples are the Lobula Giant Movement Detector (LGMD) for collision detection based on locust eyes (Blanchard et al., 2000) and flying motion detectors. (2) Neuromorphic chips based on fly eyes (Harrison, 2000) and VLSI retinal circuits (Liu and Kramer et al., 2002), and (3) Bio-inspired behavioral strategies (Srinivasan et al 2001). In these insect vision-inspired designs, the goal has been to make fast, robust, lightweight and low-power vision systems. Another feature is that analog-VLSI has been the dominant choice in insect-vision-based chips. Ruffer and Franceschini (2004) have designed neuromorphic eyes for a mini-UAV with eye weights of only 0.8g and a weight of only 100g for the entire rotorcraft. Tests reveal that these artificial vision chips (even the most flexible analog-VLSI fly eye) still have significant gaps with real insect visions systems upon which the chips are based. This indicates that a better understanding of insect eye motion detection has to be gained to make further breakthroughs (Rind, 2005). More recently, Fife and Archibald (2007) applied FPGA approach to support real-time vision processing for the small UAV.

Neural Network Modelling and Dendritic Neuronal Computing. It is interesting to note that recent advances in neural biology may change our thinking about modeling neural networks, perhaps including the ANN (Artificial Neural Network). Vogels and Rajan et al. (2005) present an excellent critical review on neural network dynamics, and they call for the models that go beyond describing and adapting to the input-output dynamics. The mathematical modeling has to address the fundamental property of the brain, that is, the neural circuits perpetually generate complex activity patterns of extraordinarily rich spatial-temporal structure, yet they remain highly sensitive to sensory inputs. London and H¨ausser (2005) offered the perspective from the computation capability of single neuron, the so-called *dendritic computation*. They argue that the computing "tool kit" of dendrites may play roles well beyond currently acknowledged properties. Ma and Krings (2007) suggested that the neuronal population in insect sensory system such as the visual neurons that support the compound eyes, the PN (Projection Neurons) in AL (Antennal Lobe), or the ORN (Olfactory Receptor Neurons may be emulated to develop a new grid computing

architecture. It seems that the integrated model of grid computing with dendritic computing may offer new insights. That is, a grid-computing infrastructure is supported by tool kits from dendritic nodes.

Biosensoring. A biosensor is an integrated device that combines a biological component with a physicochemical detector component that converts a biological response to specific substances being monitored into an electrical signal. An annual review has been published by Rich and Myszka since 1999 (Rich and Myszka, 2005).

Cellular Computing, Agent-based Computing and Swarm Intelligence. Amos et al. (2004) and Dogaru (2003) presented two of the latest review on cellular computing. Ma and Krings (2007) contrasted cellular grids in cellular computing *vs*. *neuron populations* in insect sensory systems. The latter can be the visual neurons that support the compound eyes, the PN (Projection Neurons) in AL (Antennal Lobe), or the ORN (Olfactory Receptor Neurons) from insect sensory organs (sensilla). This neuronal population can be emulated to develop a new grid computing architecture.

Agent computing is another field where insect model plays a significant inspirational role. The most well- known paradigm should be Swarm intelligence (Bonabeau et al. 1999, Dorigo and Stützle 2004), which is inspired by ants pheromone-modulated cooperative behaviors. This is in contrast with the non-cooperative behavior we discussed in subsection 2.3. Lodding's (2004) biomorphic software and the design patterns of Babaoglu and Canright et al. (2006), and Dobson and Massacci's (2006) are more examples of agent-based adaptive computing.

Molecular Networks and System Biology. A biological cell is a complex "network of networks" from the information processing perspective. Physiologically, it is an integrated device consisting of several thousands of types of interacting proteins. Molecular network is often used as a generic term to refer to the networks involved in cell biology (Alon, 2007). The gene regulatory networks and various molecular networks in cells involved extremely complex yet robust networks, which is one of the focuses of the newly emerged *system biology*. There are enormous opportunities for computer scientists to contribute and to learn from the fields. The following are a few review references: on genome project by Ideker and Galitski et al. 2001, the bio-mimetic nano-scale reactors and networks by Karlsson et al. (2004), molecular networks by Galitski (2004), gene regulatory networks by Davidson (2006).

4 PERSPECTIVE

In the following, we mention some promising research topics that seem not yet being explored. In various previous sections and Ma and Krings (2007), we briefly discussed them in corresponding context; the following is simply a list of summary statements. (1) The new grid computing architecture that emulates neuronal population such as vision neurons for insect compound eyes. This neuronal population architecture may be further integrated and enhanced with dendritic computing (2) Wireless sensor networks that emulate the insect chemosensory networks and the multimodal architecture that has several *parallel* networks concurrently in operation (such as audition, vision chemosensory, etc.). In addition, the bio-robustness mechanism in these insect sensory networks should be captured. (3) The implications of non-cooperative behaviors in social insects to swarm intelligence, evolutionary computing, and to devising new optimization algorithms. (4) Insect audition, which was considered as less developed in insects until recently, is recognized now as underestimated in entomology (Drosopoulos and Claridge, 2006). Still the field of insect audition has received little attention from the bio-inspired perspective. It is interesting to note that insects audition truly resembles the engineered wireless communications. (5) The integration of technologies that are developed for sensors, robots, MAVs, and neuromorphic technologies, in particular, the multi-modality integration may provide better solutions for the sophisticated MAV flights control, especially in unstable and hostile military operations environments.

In the recent report on "Computing and Biology" from the US National Academy of Sciences (National Research Council, 2005), the ants colony optimization and neural-inspired sensors, together with hundreds of other research topics, were recommended as fields of strategic scientific and technological significances. However, the majority of topics on insect sensory systems, such as those discussed in this series of articles, were omitted. There was significant coverage (nearly 4 pages) on ants colony optimization in the National Academies' report. This coverage may also indicate the significance of the areas omitted in the report, which, in our opinion, should prove to be as promising as the ants colony optimizations, if not

more.

ACKNOWLEDGEMENTS

This research was partially supported by a *UAV and Ad hoc Networking* research grant from the US DOE INL. We wish to thank the two anonymous reviewers for their insightful comments.

REFERENCES

Alon, U. 2007. *An Introduction to Systems Biology*: *Design Principles of Biological Circuits*. Chapman.

Amos, M. 2004 (ed.). *Cellular Computing*. Oxford University Press. 217pp.

Babaoglu, O. and G. Canright et al. 2006. *ACM Trans. on Auton. & Adap. Syst.* 1(1):26–66.

Blanchard, M. and F. C. Rind, P. F. M. Vershure. 2000. *J. Robot. Auto. Syst*. 30:16-38.

Bonabeau, Eric. M. Dorigo, G. Theraulaz. 1999. *Swarm Intelligence*. Oxford University Press. 320pp.

Christensen, T. A. (ed). 2005. *Methods in Insect Sensory Neuroscience*. CRC Press. 435pp.

Davidson, E. H. 2006. *The regulatory genome: gene regulatory networks in development and evolution*. Academic Press. 304pp.

Detrain, C., J. L. Deneubourg, and J. M. Pasteels.(eds). 1999. *Information Processing in Social Insects*. Birkhauser Verlag. 415pp.

Dobson, S., F. Massacci, et al. 2006. *ACM Transactions on Autonomous and Adaptive Systems*, 1(2): 223–259

Dogaru, R. 2003. *Universality and Emergent Computation in Cellular Neural Networks*. World Scientific Press.

Dorigo, M., T. Stützle. 2004. *Ants Colony Optimization*. The MIT Press.

Fife, W. S. and J. K. Archibald. 2007. *EURASIP J. on Embedded Systems*. (1) 33

Dressler, F. I. Carreras. 2007. *Advances in Biologically Inspired Information Systems: Models, Methods, and Tools*. Springer, 302pp.

Drosopoulos, S. and M. F. Claridge. (eds). 2006. *Insect Sounds and Communication.* Taylor & Francis. 532pp.

Galitski, T. 2004. Molecular networks in model systems. *Annu. Rev. Genomics Hum. Genet*. 2004. 5:177–87

Gullan, P. J. and P. S. Cranston. 2005. *The insects: an outline of entomology*. 3rd Ed. Blackwell Publishing.

Harrison, R. R. 2000. *An analog VLSI motion sensor based on the fly visual system*. Ph.D. Dissertation, CalTech, CA.

Ideker, T., T. Galitski1, and L. Hood. 2001. *Annu. Rev. Genomics Hum. Genet*. 2001. 2:343–72

Karlsson, M., M. Davidson et al. 2004. *Annu. Rev. Phys. Chem*. 2004. 55:613–49

Koickal, T. J., A. Hamilton, S. L. Tan, J. A. et al. 2007. *IEEE Trans. Circuits and Systems.*54(1): 60-73.

Klowden, M. J. 2007. *Physiological Systems in Insects*. 2nd ed. Academic Press.

Liu, S. C., J. Kramer, G. Indiveri, T. Delbrück, R. Douglas2002. *Analog VLSI: Circuits and Principles*. MIT Press. 472pp.

Lobontui, N, Goldfarb, M., et al. 1999. Design and Analysis of an Elastodynamic Inch-Worm Robotic Insect. *IEEE Conference on Robotics and Automation*, pp. 2120-2125, May 1999.

Lodding, K. N. 2004. *Queue,* June 2004:68-75.

London, M. and M. H¨ausser 2005. Dendritic Computation. *Annu. Rev. Neurosci*. 2005. 28:503–32

Michelson, R.2002. The Entomopter. "*Neurotechnology for Biomimetic Robots*", The MIT Press, pp.481-50

Ma, Z., and A. W. Krings. 2007. Insect Sensory Systems Inspired Computing and Communications. Technical Reports, TR-CS-09-01-2007, Computer Science Dept. University of Idaho.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press.

Miesenbock, G. and I. G. Kevrekidis. 2005. *Annu. Rev. Neurosci*. 2005. 28:533–63

Motamed, M., and J. Yan. 2005. A review of biological, biomimetic and miniature force sensing for microflight. International Conference on Intelligent Robots and Systems (IROS). 2-6 Page(s): 3939 - 3946.

National Research Council of National Academies. 2005. Catalyzing Inquiry at the Interface of Computing and Biology. The National Academy Press.

Pornsin-sirirak, T. N. Y. C. Taia, et al. 2001. *Sensors and Actuators* A 89. (2001): 95-103

Ratnieks, F. L. W., K. R. Foster, and T. Wenseleers. 2006. Conflict resolution in insect societies. *Annu. Rev. Entomol*. 2006. 51:581–608

Ruffer, F. and N. Franceschini, 2004. "Visually guided micro-aerial vehicle: automatic take off, terrain following, landing and wind reaction," in P*roceedings of ICRA 2004)*, vol. 3, pp. 2339-2346.

Rich, R. L. and D. G. Myszka. 2005. *J. Mol. Recognit*. 18: 431–478

Rind, F. C. 2005. Bioinspired sensors. in. "*Methods in Insect Sensory Neuroscience*". ed. by T. A. Christensen. CRC Press. pp. 213-235.

Srinivasan. 2003. *Journal of Robotic Systems* 20(1), 35–42

Srinivasan, M. V., S. W. Zhang. et al. 2001. *Biol. Bull*. 200:216-221.

Steltz, E., S. Avadhanula, R.J. Wood, and R. S. Fearing. 2005. Characterization of the Micromechanical Flying Insect by Optical Position Sensing, *IEEE Int. Conf. on Robotics and Automation*, Barcelona, April 2005.

Stocker, A. A. 2006. *Analog VLSI Circuits for the Perception of Visual* Motion. Wiley.

Thakoor, S., and N. Cabrol, et al. 2003. *J. of Rob. Syst.* 20(12), 687–706.

Vogels, T. P., K. Rajan, and L. F. Abbott. 2005. Neural Network Dynamics. *Annu. Rev. Neurosci*. 28:357–76

Wood, R. J., S. Avadhanula, E. Steltz, M. Seeman, et. al. 2005. Design, Fabrication and Initial Results of a 2g Autonomous Glider. *The 31st Annu. Conf., IEEE Indust. Electron. Soc.*, Raleigh North Carolina.

DIFFERENCES IN PHYSIOLOGICAL RESPONSES TO THE INTENSITY OF MENTAL STRESS

Chi'e Soga, Chikamune Wada
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
soga-chie@edu.life.kyutech.ac.jp, wada@ life.kyutech.ac.jp

Shinji Miyake
University of Occupational and Environmental Health, Japan
1-1 Iseigaoka, Yahatanishi-ku, Kitakyushu, Fukuoka, 807-8555, Japan
myk@health.uoeh-u.ac.jp

Keywords: Autonomic nervous system, RRI, plethysmogram, skin potential level, blood pressure, mental workload.

Abstract: It is widely understood that mental stress produces various physiological changes. Though the relationship between mental stress and physiological response has been extensively reported, few reports have tried to clarify the relationships between various physiological responses and the intensity level of stress. In this study, we investigated autonomic nervous system activities to find a physiological index based on which we can evaluate the intensity of mental stress. As a result, we found that there were different response patterns for each physiological index. We consider that each physiological index shows different feelings and/or situations related to mental stress.

1 INTRODUCTION

In our country, the increase in psychiatric disorders, such as depression and schizophrenia, is noted. The number of suicides per year has been steadily high in recent years, with more than 30,000 people a year since 1998. This increase in psychiatric disorders and high rate of suicide are serious problem in Japan.

It is believed that these daily stresses play a role in a number of psychiatric disorders. If we can evaluate daily stress quantitatively and determine our own or other people mental state, this could contribute to the prevention of various diseases caused by mental stress.

It is widely understood that physiological changes induced by mental stress are related to the autonomic nervous system, and can affect the heart rate, blood pressure and plethysmogram. The relationship between mental stress and physiological feedback has been extensively reported (Takatsu et al., 2000, Mishima, Kubota and Nagata, 1999).

We also consider that establishing a quantitative evaluation method for mental stress will help prevent diseases caused by mental stress. It is necessary to examine the intensity of stress to realize a quantitative evaluation of mental stress. However, few reports have tried to clarify the relationships between physiological responses and the intensity of stress. In this paper, we investigated the autonomic nervous system activity in the three conditions for the intensity of mental stress.

2 METHOD

We used a mental arithmetic task as the mental workload and measured the physiological and subjective responses.

2.1 Task

A target three-digit number, several two-digit numbers and an OK button were displayed on a computer screen. The participants were required to select the combination of three two-digit numbers whose sum is equal to the target number and click the OK button. Whenever a participant clicked the

OK button, the sum of the selected numbers was shown on the screen. If the sum was equal to the target number, the next arithmetic question was displayed. The participant could choose different combinations of numbers until the right one was chosen. The elapsed time and the number of correct answers were also displayed on the screen. The task screen is shown in Figure 1.

The three conditions were used for the intensity of mental stress. The conditions were as follows:

TASK1 (High level): The number of two-digit number is ten.
TASK2 (Medium level): The number of two-digit number is eight.
TASK3 (Low level): The number of two-digit number is five.

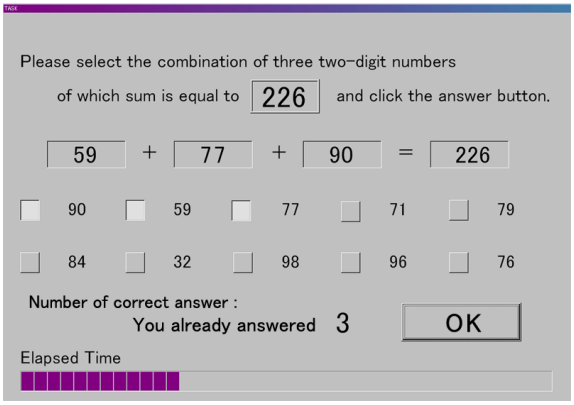


Figure 1: Mental arithmetic task screen of high level.

2.2 Physiological Measurements and Subjective Assessment

We measured the electrocardiogram (ECG), plethysmogram (PTG), blood pressure (BP), tissue blood pressure (TBV) and skin potential levels (SPL). These signals were recorded in a PC at a 1-kHz sampling rate. Also, stroke volume (SV) and cardiac output (CO) were obtained every heartbeat.

The R-R interval (RRI), LF/HF ratio, systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP=DBP+(SBP-DBP)/3), baroreceptor reflex sensitivity (BRS=square root of (LF of SBP/LF of RRI)), amplitude of the PTG and total peripheral resistance (TPR=MBP/CO) were calculated.

The National Aeronautics and Space Administration Task Load Index (NASA-TLX), Profile of Mood States (POMS) and the semantic differential method (SD method) were used to obtain the subjective responses.

The NASA-TLX is a widely used subjective workload assessment technique (Hart and Staveland, 1988). The WWL value was calculated. The POMS consists of 65 adjectives and assesses six mood states dimensions. We used 24 adjectives related to Tension-Anxiety (TA), Vigor (V) and Fatigue (F) for reducing the participant's burden. The SD method comprising seventeen items was used to assess the participants' emotions.

2.3 Procedure

The participants were familiarized with what to expect during the task before the start of the experiment. Each participant underwent the experiment procedure once.

The experimental procedure was as follows:

1. Rest1 (6 min: PRE1)
2. High level arithmetic (6 min: TASK1)
3. POMS, NASA-TLX and SD method
4. Rest2 (6 min: PRE2)
5. Medium level arithmetic (6 min: TASK2)
6. POMS, NASA-TLX and SD method
7. Rest3 (6 min: PRE3)
8. Low level arithmetic (6 min: TASK3)
9. POMS, NASA-TLX and SD method
10. Rest4 (6 min: POST)

The participants were instructed to provide at least fifteen correct answers in six minutes and not to give up until the six-minute run was finished. If the participants completed the task (i.e. gave the minimum number of correct answers) before the deadline, they were required to continue giving correct answers. After six minutes, the run was finished, and the participants were asked to cease the mental arithmetic activity even if they had not completed the task. The number of correct answers was displayed on the screen to let the participants know when the task was completed.

2.4 Participants

Sixteen healthy male graduate students aged 21 to 32 (average: 23.8 yrs.) participated in this study. All participants gave their written informed consent.

3 DATA ANALYSIS

3.1 Statistical Analysis

The data were divided into seven 6-minute blocks (PRE1, TASK1, PRE2, TASK2, PRE3, TASK3 and POST). All parameters were standardized for each participant.

The results were analyzed by repeated measures of analysis of variance (ANOVA) using SPSS 11.0J. The degree of freedom was adjusted using the Greenhouse-Geisser correction. Tukey's honestly significant difference (Tukey's HSD) test was used in the *post-hoc* analysis.

3.2 Subjective Assessment

Three factors (FACTOR1, FACTOR2 and FACTOR3) were obtained from the seventeen items of the SD method using Factor Analysis. The factor scores, NASA-TLX scores, POMS scores and all physiological indices were standardized for each participant. The correlation coefficients of 48 pairs of subjective assessments and physiological indices were calculated.

4 RESULTS AND DISCUSSION

4.1 Differences in Physiological Response

The WWL value of Task3 was significantly low ($p<0.05$). The results of the multiple comparison revealed that there were no significant differences in WWL value between Task1 and Task2. This result indicated that the condition of Task3 was simpler than the conditions of the other two tasks.

There were two patterns for each physiological response in both different cases: task periods and the resting periods. Figure 2 shows the averages of the SPL and the amplitude of the PTG for all the participants. Both parameters were significantly lower in TASK1 and TASK2 comparing with resting periods. The SPL values indicated that the changes during the task were smaller when the difficulty level became low. The RRI, SBP, DBP, BRS and TPR showed a tendency similar to that of the SPL. The amplitude of the PTG indicated that the responses after the completion of the task (during rest) were smaller when the difficulty level became low. The TBV, CO and SV showed a tendency similar to the amplitude of the PTG. There was no discernible tendency in the LF/HF ratio.

In a previous study, we found that there were different recovery patterns for each physiological index (Soga, Miyake and Wada, 2007). In this study, there were also different recovery patterns for the various physiological indices. These results suggest that each physiological response corresponds to different feelings and/or situations. A difference in response during the task might correspond to "Executing task" and the feeling of "Tension." A difference in response after the task might correspond to "Task performance" and the feeling of "Regret."

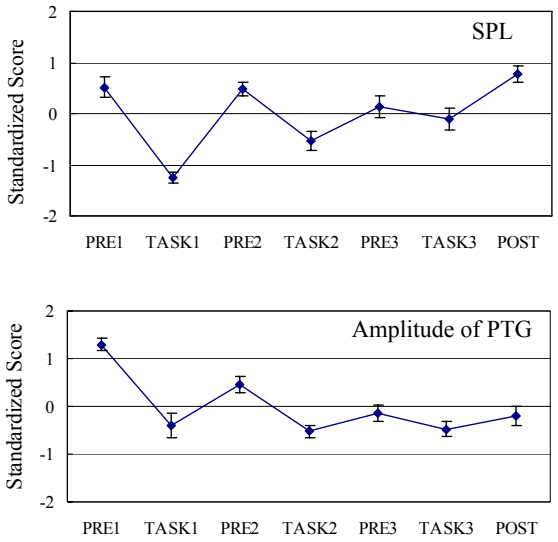


Figure 2: The average changes in the physiological responses for each block. The bars indicated the standard errors of the mean.

4.2 Correlation Coefficient between the Subjective Measurements and the Physiological Indices

There were significant correlation between the FACTOR1 score and the DBP ($r=-0.340$; $p<0.05$), MBP ($r=-0.312$; $p<0.05$) and SPL ($r=0.506$; $p<0.001$). There were significant correlation between the FACTOR2 score and the SPL ($r=0.322$; $p<0.05$). The FACTOR3 score significantly related to the SBP ($r=-0.350$; $p<0.05$), DBP ($r=-0.378$; $p<0.05$), MBP ($r=-0.364$; $p<0.05$), BRS ($r=0.415$; $p<0.01$), RRI ($r=0.433$; $p<0.01$) and SPL ($r=0.312$; $p<0.05$). The FACTOR1 consisted of the following items: "Difficult-Easy," "Troublesome-Smooth," "I felt the task duration was short-I felt the task duration was long," etc. Therefore we consider that

the FACTOR1 relates to the degree of difficulty of the task. The FACTOR2 consisted of the following items: "Dislike-Like," "Boring-Interesting" and "I want to execute the task-I don't want to execute the task." The FACTOR3 was as follows: "Respond randomly-Respond after proper calculation." We consider that the FACTOR2 relates to the participants' concentration and the FACTOR3 relates to the participants' attitude.

The NASA-TLX revealed that there were significant correlation between the TD and SPL ($r=-0.676$; $p<0.001$), and the OP and the amplitude of the PTG ($r=-0.324$; $p<0.05$). The POMS showed that there were significant correlation between V and the SPL ($r=-0.381$; $p<0.05$), and F and the amplitude of the PTG ($r=-0.308$; $p<0.05$).

Although some physiological indices seemed to correlate with the subjective assessments because the correlation values were small, there were only three blocks for standardization, and all the data were pooled. At least, we consider that each physiological response corresponds to different feelings and/or situations. This suggests that the classification of physiological responses according to the results of the subjective assessment is helpful in investigating the complex information contained in each physiological index. In our past study, we found that the SPL related to the time pressure (Soga, Miyake and Wada, 2007). In this study, the SPL results showed a tendency similar to that found in our past study. Therefore we consider that the SPL is a sensitive index for the estimation of mental stress.

5 CONCLUSIONS

We found that there were two patterns for each physiological response of the autonomic nervous system during the task periods and the resting periods at three different difficulty levels (high, medium and low). Significant difference in PTG amplitude between task and after task resting period was disappeared in the last two blocks (TASK3 and POST). In addition, we found that there were significant correlation between the physiological changes and the subjective assessments.

These results suggest that each physiological response corresponds to different feelings and/or situations related to mental stress. Further experiment should be done to confirm this result. Our final aim is to establish a quantitative evaluation method for mental stress.

REFERENCES

Hart S H. and Staveland L. E., 1988. *Elsevier Science Publishers B.B., pp.139-183, Amsterdam Development of NASA-TLX (Task Load Index): Result of Empirical and Theoretical Research, In P. A. Hancock and N. Meshkati (eds.), Human Mental Workload.*

Mishima N, Kubota S and Nagata S, 1999. *Psychotherapy and Psychosomatics, Vol.68, No.4, 207-213, Psychological correlates of relaxation induced by standard autogenic training.*

Soga C, Miyake S, Wada C, 2007. *SICE Annual Conference 2007, 1366-1371, Recovery patterns in the physiological responses of the autonomic nervous system induced by mental workload.*

Takatsu H, Munakata M, Ozaki O, Yokoyama K, Watanabe Y and Tanaka K, 2000. *T.IEEE Japan, Vol.120-C, No.1, 104-110, An evaluation of the quantitative relationship between the subjective stress value and heart rate variability (in Japanese).*

AN FPGA PLATFORM FOR REAL-TIME SIMULATION OF TISSUE DEFORMATION

Samson Ajagunmo and Aleksandar Jeremic
Department of Electrical and Computer Engineering, Main St., Hamilton, Ontario, Canada
ajaguns@mcmaster.ca, jeremic@ece.mcmaster.ca

Keywords: Reconfigurable Architecture, Tissue deformation, Matrix-by-Vector Multiplication, Conjugate Gradient Method, Field Programmable Gate Arrays.

Abstract: The simulation of soft tissue deformations has many practical uses in the medical field such as diagnosing medical conditions, training medical professionals and surgical planning. While there are many good computational models that are used in these simulations, carrying out the simulations is time consuming especially for large systems. In order to improve the performance of these simulators, field-programmable-gate-arrays (FPGA) based accelerators for carrying out Matrix-by-Vector multiplications (MVM), the core operation required for simulation, have been proposed recently. A better approach, yet, is to implement a full accelerator for carrying out all operations required for simulation on FPGA. In this paper we propose an FPGA accelerator tested for simulating soft-tissue deformation using finite-difference approximation of elastodynamics equations and conjugate-gradient inversion of sparse matrices. The resource and timing requirements show that this approach can achieve speeds capable of carrying out real-time simulation.

1 INTRODUCTION

Some of the most common procedures in clinical practice (e.g. the insertion of subcutaneous needles in the tissue for biopsy of deep-seated tumors) are extremely sensitive to guiding algorithms and initial placement of the needle. One of the current trends in this field is the development of virtual simulators for tissue deformation. Realistic simulation of tissue deformation undergoing needle insertion is the bottleneck of all virtual simulators.

The deformation of soft tissue is determined by elastodynamic partial differential equations (PDEs) (Fung, 1987), defined over irregular domains (human organs). A solution to these PDEs cannot be obtained analytically due to their nonlinearity and irregular shape of the domain. In order to solve these equations we need one of commonly used discretizaion techniques: the finite-difference method (FDM) and the finite-element method (FEM). In both methods, the domain of interest is discretized and the corresponding PDEs are transformed into linear equations. The resulting linear system is then solved using numerical methods such as Newtons method, conjugate-gradient method (CGM) etc.

Most of the resent work done in this area focused on speeding up numerical methods by implementing efficient matrix-by-vector multiplier units (MVU) on FPGA. In (Ramachandran, 1998) the author investigated the performance effects of using an FPGA based MVU to carry out an MVM. The MVU was able to achieve a performance of 36 MFLOPS with a matrix generated using the Finite-Element method. In (Zhuo et. Al, 2005) the authors also developed an MVU for MVMs that involved sparse matrices. Their method involved using only the non-zero elements of a matrix to carry an MVM. The design in (Zhuo et al, 2005) attained a performance of 350 MFLOPS for all their test cases. This is a 900% increase in performance when compared with results in (Ramachandran, 1998). Note however, that as of 2005, FPGAs were capable of higher clock frequencies than in 1998, which most likely was one of rather important factors for such improvement.

In this paper we propose an FPGA platform for real-time simulation of tissue deformation using FDM model and CGM for solving the corresponding linear system. We will implement the CGM, a full numerical method, in hardware on an FPGA. We will also exploit the fact that the "stiffness" matrix is sparse and band-limited. Our preliminary results indicate that we can achieve sufficiently high computational rate even with larger size meshes.

2 BACKGROUND

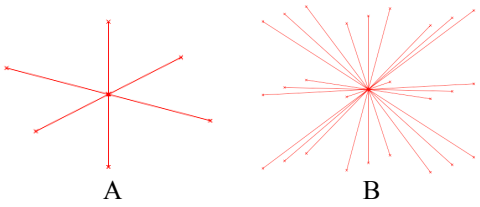


Figure 1: Connection Pattern Models.

We model the soft tissue as a three-dimensional grid of uniformly distributed nodes (material points) connected together by springs that model the elastic properties of the tissue. To model the connection between two material points we use two connection patterns shown in Figure 1. Using a quasi-static approach when an external force acts on a certain node, it causes a change in the length of the springs connected to that node. This also creates opposing forces in these springs so as to keep the system of connected springs in equilibrium. This relationship, for a given direction d at a node i,j,k, is given in (1) by the function $f^d_{i,j,k}$.

$$f^d_{i,j,k} = \sum_{n=-1}^{1}\sum_{m=-1}^{1}\sum_{l=-1}^{1} k^d u^d_{i+l,j+m,k+n} \qquad (1)$$

Assembling these nodal equations for every node yields a set of linear simultaneous equation that describes the system in direction d. These equations can be represented in matrix form as shown in (2), where d, K^d, f^d are the displacement vector,

$$f^d = K^d d \qquad (2)$$

characteristic ("stiffness") matrix, and load vector respectively, in the direction of d. To solve the equation in (2), for each direction d, we utilize the CGM, which is an iterative technique that can be carried out amenably on FPGA at speeds capable of real-time simulations.

3 CGM ACCELERATION

The CGM consists of a series of one or more MVM and vector-by-vector multiplication (VVM). Since MVMs are more computationally intensive than VVMs, the effective bottleneck of this numerical method are the MVMs. The acceleration of the CGM involves designing hardware optimised for carrying out operations needed by the CGM (CGM Accelerator), and the speeding up of MVMs. Speeding up MVMs involves dividing the multiplying matrix **K** and vector **v** into smaller appropriately dimensioned sub-matrices and sub-vectors. Each of these sub-matrices and sub-vectors are then used by a series of MVUs working in parallel, to carry out the required MVM. Each of these sub-matrices must be stored in separate memory blocks, one for each of the MVUs that will be working in parallel.

The CGM accelerator consists of a series of MVU for carrying out MVMs, and a Scalar-Vector Unit (SVU) for carrying out the remaining scalar and vector operations in the CGM.

3.1 SVU Design

As mentioned earlier, the SVU carries out all the required operations in the CGM except for the MVM. In Figure 2, we show the set-up that carries out these operations. Most of the operations in the CGMs main loop (shown below) are dependent on each other hence; they must be carried out sequentially in the order of dependence. For example α must be updated before x or r is updated, and r must be updated before β is updated. The updating of x and r are, however, independent of one another, so they can be carried out simultaneously. However, the amount of time, one clock cycle, that is saved is not justified when considering that the

$n = 0$	
$r_n = b - Kx_0$	Residual error based on initial guess $x = x_0$
$p_n = r_n$	Initial Direction
for $n = 1 : N$	
$n = n + 1$	
$\alpha = r^T_{n-1} r_{n-1} / p^T_{n-1} K p_{n-1}$	Calculate New Step Size for x and r
$x_n = x_{n-1} + \alpha * p_{n-1}$	Calculate New x
$r_n = r_{n-1} - \alpha * K p_{n-1}$	Calculate New Residual r
$\beta = r^T_n r_n / r^T_{n-1} r_{n-1}$	Calculate New Step Size for p
$p_n = r_n + \beta * p_{n-1}$	Calculate New Search direction p
end	
$x = x_n$	Final Result

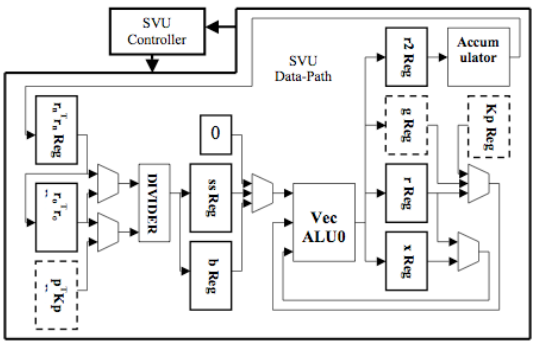


Figure 2: SVU Design.

amount of resources that is required to update x and r will be doubled. However, we can double the size of the system that the SVU can handle by allowing an extra clock cycle. The completion time for the SVU is always fixed, unlike the MVU were the completion time will vary with the size of the matrix that it uses for the MVM.

As seen in Figure 2, the three main modules used in the SVU data-path are a Divider, a Vector ALU (VecALU), and an Accumulator. The operations performed by these modules are described next.

Divider: This module is used to calculate α and β, which are used by VecALU.

VecALU: This is an arithmetic logic unit (ALU) that specifically carries out vector-vector or vector-scalar operations. The residual r, search direction g, and deformation x are updated here. The module uses previous values along with α and β to generate new values. The new value of r is passed to the Accumulator.

Accumulator: This module basically sums the elements of the register r2 reg. The result of this summation is the 2-norm of vector r. Hence, each element of register r2 Reg is the square of the corresponding element in r. The divider uses this 2-norm value in the calculation of α and β.

The SVU-Control controls the flow of information among the registers and modules in the SVUs data-path. As seen in Figure 2, there are three registers, shown by dashed lines, one for the multiplying vector, while the others are for the MVU results. These are the three registers used to pass information between the SVU and the MVU. The multiplying vector register **g Reg** is used for passing the direction vector to the MVU, while the result registers, **p^TKp Reg and Kp Reg,** are used for receiving the MVU results (p^TKp and Kp).

3.2 MVU Design

This MVU is designed specifically for MVMs, of the form Kp and p^TKp, which may involve sparse matrices. The design, shown in Figure 3, requires only the non-zero elements of the matrix to be stored in the memory. The non-zero elements are stored in memory as part of a simple 32-bit instruction format, shown below, that was designed for the MVU. Further, these non-zero elements are stored in memory using fixed-point format.

a(1bit)	*b*(1bit)	*c*(9bits)	*d*(21bits)

a *1st bit determine s end of matrix.*
b *2nd bit determine s end of row.*
c *3rd to 11th bits used to determine the column of the nonzero value.*
d *last 21 bits give the nonzero value.*

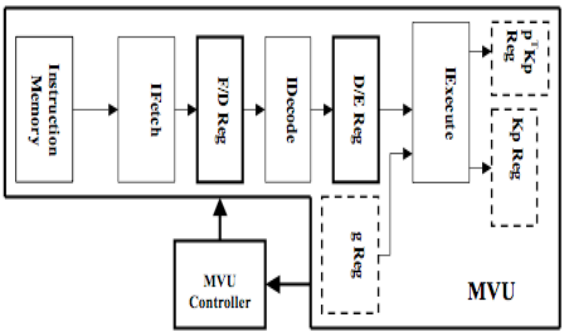


Figure 3: MVU Design.

The MVU data-path is pipelined and divided into three modules, namely, Instruction Fetch module (IFetch), Instruction Decode module (IDecode), and Execute module (IExecute).

IFetch: This module just fetch's the next instruction from memory and forwards it to IDecode for use. The instructions are read sequentially with the addresses gotten from a sequential counter.

IDecode: The instruction is decoded here using the format described earlier. It is determined here if the end of the current row or column (ERC) or the end of matrix (EM) has been reached. The address of the next vector element needed for the next multiplication is also determined here.

IExecute: This module basically performs the traditional MVM (i.e. taking the inner product of each row and the multiplying vector, starting with the first row) using a set of multipliers and accumulators. The calculation of p^TKp and Kp are done concurrently, with the appropriate values stored in the appropriate result registers.

The MVU-Controller controls the flow of information among the registers and modules in the MVU data-path. As discussed earlier, the MVU result registers, and multiplying vector register are used for passing information between the SVU and MVU.

4 RESOURCE USAGE AND PERFORMANCE

FPGAs contain three main resources namely, multipliers, logic elements and registers. Of these three, the multipliers are of least abundance. This makes them the bottleneck of any design for applications that are heavily dependent on the usage of multipliers. For this reason, we use the multiplier usage as the primary measure of our designs resource usage, as it is the deciding factor in the maximum size of the system that can be solved on

one FPGA. Figure 4 shows the multiplier usage of our CGM accelerator implementation for different number of MVUs and problem sizes n (number of nodes). We implemented the CGM accelerator on Altera's DE2 development board using the Quartus II development software. The implementation can be clocked at speeds up to 133MHz.

The completion time for one iteration of the CGM is given by (5), the sum of the completion times for the SVU and MVU. Of these two, T_{MVU}

$$T = T_{MVU} + T_{SVU} \tag{3}$$

is the only time that can be improved on by using the technique described in section 3. Minimizing T_{MVU} effectively reduces the to time to carry out the CGM. Hence T is a good measure of performance for our CGM accelerator. We used a two-pronged approach to test for the timing performance of the CGM accelerator. Firstly, we used Quartus II simulator to get preliminary test results for the CGM accelerator. Secondly, we will verify these simulation results with test results from the hardware implementation of the CGM accelerator. These tests are done at 100MHz. In Figure 5 we show the preliminary results for the computation time, T, of one iteration of the CGM as a function of number of MVUs and problem size n.

MFLOPS, given by (4), is another common measure of performance. MFLOPS is a measure of the number of floating point operations per second. n is the size of the problem and m is the of average number of nonzero elements per row. In Figure 6, we show the MFLOPS performance as a function of problem size for systems generated using connection pattern B in Figure 1. Our CGM accelerator was

$$MFLOPS = \frac{Total\#of\ flops/iter.}{compute\ time/iter.} = \frac{2mn+3n+2}{T} \tag{4}$$

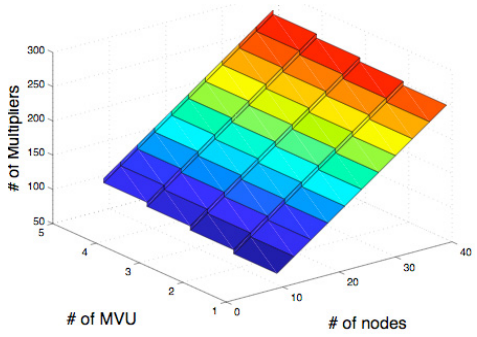


Figure 4: Multiplier Usage.

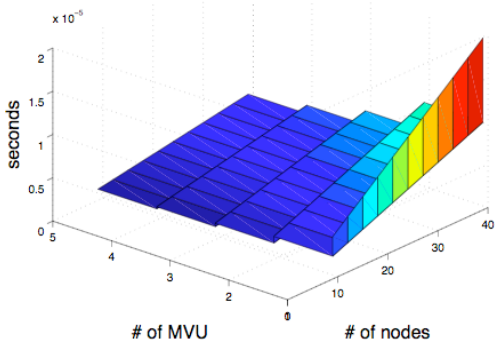


Figure 5: Computation time.

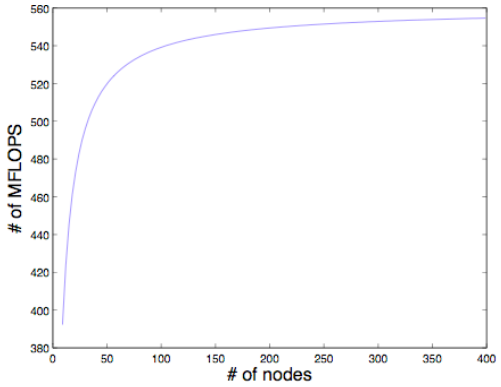


Figure 6: MFLOPS Performance.

able to achieve more than 540 MFLOPS with 5 MVUs working in parallel. As you can see in Figure 6, the performance of the system plateaus as n gets larger. This is mainly due to the fact that the number of MVUs is fixed. However, for better performance, we can use more MVUs in parallel. Note, however, that the use of more than one MVU in parallel means that fewer multipliers are available for use by the SVU, as the number of multipliers available on the FPGA is fixed. Hence, the amount of resources available determines the optimal number of MVUs that can be used in parallel, and size of problems that can be solved.

5 CONCLUSIONS

We proposed and implemented an FPGA based CGM accelerator for carrying out real-time simulation of tissue deformation. Our design does not require any information on the sparsity of the stiffness matrix. Further more, we gave a brief discussion on improving the speed of MVMs using parallel computing. We then looked at the resource requirements and the performance of the CGM accelerator. Our preliminary performance results

show that developing FPGA accelerators for use in real-time simulation is feasible. Our next step is to verify these results as described in section 4.

REFERENCES

Chapra, C. S., Canale, P. R., 2002. *Numerical Methods for Engineers*, McGraw Hill. NewYork, NY, 4th edition.

DiMaio, S. P., Salcudean , S. E., 2002. *Needle Insertion Modelling for the Interactive Simulation of Percutaneous Procedures*. In the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II. Springer-Verlag 253-260.

Fung, Y.C., 1987, Biomechanics, Springer-Verlag, New York.

Goulb, H. G., Van Loan, F. C., 1996. *Matrix Computations*, The John Hopkins University Press. London, 3rd edition.

He, C., Qin, G., Zhao, W. *FPGA-Based High-Order Finite Difference Method for Linear Wave Modelling Problems*. Retrieved June 11, 2006, from http://lacsi.rice.edu/symposium/symposumdownloads.

Hennessy, L. J., Patterson, A. D., 2003. *Computer Architecture: A Quatitative Approach*, Morgan Kaufmann Publishers. San Francisco, CA, 3rd edition.

Ramachandran, K., 1998. *Unstructured Finite Element Computations on Configurable Computers*. Blacksburg, Virginia: University Libraries, Virginia Polytechnic Institute and State University.

Rocha, K. M. C. *Numerical Techniques for Real Options.* Retrieved May 15, 2006, from
http://www.puc-rio.br/marco.ind/katia-num.html.

Zhuo, L., Prasanna, V. K., 2005. *Sparse Matrix-Vector Multiplication on FPGAs*. In Computation algorithms for FPGA, ACM/SIGDA 13th international symposium on Field programmable gate arrays. ACM Press. 63-74.

Altera DE2 Development Board
http://www.altera.com/education/univ/materials/boards/unv-de2-board.html.

RELATIONSHIP BETWEEN THERMAL PERCEPTION AND MECHANICAL CHARACTERISTICS ON A PALM

Aiming at Developing a Communication Support Device for the Deaf-Blind

Chikamune Wada, Kuranosuke Sako and Hiroshi Horio
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Hibikino 2-4 Wakamatsu, Kitakyushu, Japan
wada@life.kyutech.ac.jp, sako-kuranosuke@edu.life.kyutech.ac.jp, horio-hiroshi@edu.life.kyutech.ac.jp

Keywords: Deaf-blind, tactile display, mechanical characteristics, thermal perception.

Abstract: Our final goal is to develop a portable display which will enable the deaf-blind to character on the palm through the use of tactile sensations. We propose the use of thermal stimulation as the tactile sensation, because in this way small-sized and lightweight devices can be developed. However, it might still be impossible to capture continuous movement, which is necessary to recreate characters on the palm. In past research, we found that thermal perception is dependent on the palm position. Therefore, in this study, we investigated the cause of this position dependence by comparing the skin's thermal perception and its mechanical characteristics.

1 INTRODUCTION

Since the deaf-blind suffer from both visual and auditory impairment, it is easy to understand the communication difficulties that arise due to this affliction.

In this research, we tried to create a device that can facilitate effective communication among the deaf-blind and the non-disabled.

Since schools throughout Japan teach both the deaf and the blind to write phonetic symbols known as "*kana*," most Japanese deaf-blind are familiar with this standard Japanese writing system. Then, some deaf-blind use a communication method which is writing characters on the palm of their hand using a finger. Therefore, we thought this could provide the basis for a communication device that can be used among deaf-blind.

Generally speaking, a person's finger moves continuously when a *kana* is written on the palm. Needless to say, it will be easy to transmit a shape of *kana* by using a kind of pin tactile display for the blind (Itoh, Sakai and Sakajiri, 2003). However, we think the stroke order of writing *kana* is important to let the deaf-blind know *kana*. Then, our goal is to develop a portable display which enables a person to write *kana* on the palm through the use of tactile sensations (Wada and Wada, 2003.). Although this finger movement can be reproduced by using a XY-stage, it is not possible to carry such a stage because of its size and weight. To resolve this drawback, we surmised that a portable display can be realized by making use of a Peltier element, which is both small and light. However, even if the elements were arranged, it might still be impossible to recreate continuous movement of tactile stimulation on the palm. Instead, we supposed that continuous movement can be realized using the apparent motion phenomenon.

In previous research (Horio and Wada, 2005), we investigated the optimal condition under which thermal stimulation of a Peltier element causes apparent motion. However, some subjects could not perceive the apparent motion. We hypothesized that this was due to the individual differences in the characteristics of thermal stimulation reception. Incidentally, it was difficult to measure the thermal stimulation reception while it was easy to measure mechanical characteristics of skin. Therefore, we would like to make a model by which thermal stimulation perception will be able to be estimated by mechanical characteristics.

In this study, we chose response time to cold sensation and mechanical impedance as a parameter of thermal stimulation perception and mechanical characteristics, respectively. Then, we investigated a relationship between the response time and mechanical impedance.

2 MEASUREMENT OF RESPONSE TIME

2.1 Experimental Setup

Figure 1 illustrates the outline of our experimental setup. Peltier elements (8.3 mm*8.3 mm*2.4 mm) were used to induce thermal stimulation. These elements were connected to a computer through a D/A converter and an amplifier. The thermal stimulation was controlled by the computer. Two thermocouples were used to measure the temperature: one was attached to the Peltier element to measure its temperature, while the other was attached to the palm to measure the skin surface temperature. The thermal data from the thermocouples was directly inputted into the computer through an A/D converter.

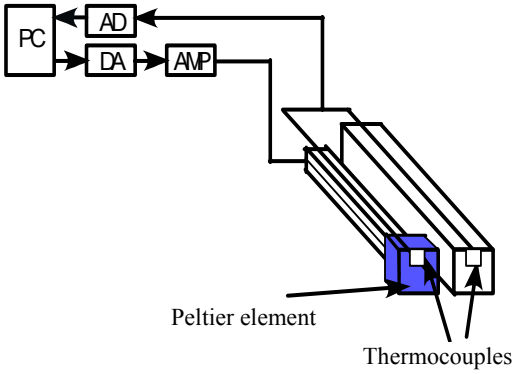


Figure 1: Experimental setup.

2.2 Experimental Procedure

The right palm was thermally stimulated. The palm length and hand breadth were measured, and the palm was divided into 16 parts (Fig. 2). The assigned number and alphabet indicates the place where thermal stimulation was induced. Figure 3 shows the thermal stimulation pattern. The vertical axis indicates the temperature of the Peltier element, while the horizontal axis indicates the elapsed time. Before the start of the experiment, the temperature of the Peltier element was adjusted to the same temperature as the subject's skin surface. The environment temperature was between 25 and 28 degrees Celsius. The subjects were six males, 22 to 25 years of age. The trial was repeated 10 times for each subject. The subjects wore earplugs and eyeshades in order not to hear environmental noise and see.

When the Peltier element surface reached the same temperature as the palmar skin surface, the element was placed on the palm. After a while, the temperature of the Peltier element decreased. The ratio of temperature decrease of the Peltier element was -5.5 degrees Celsius per second.

The subjects were asked to push a switch when they felt that the Peltier element had become cold ("Stop" in Fig. 3), after the temperature of the Peltier element started to decrease ("Start" in Fig. 3). The time interval between "Start" and "Stop" was measured. This time interval was named the "Response time."

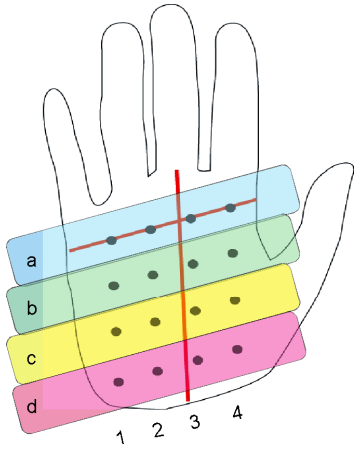


Figure 2: Stimulation points.

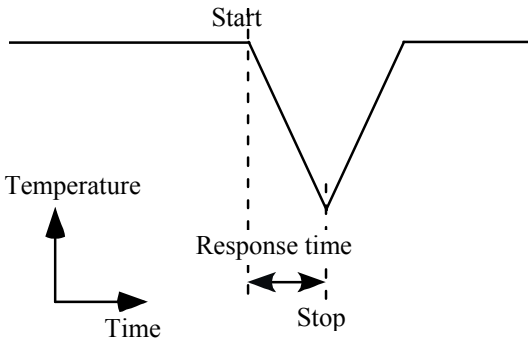


Figure 3: Stimulation pattern.

2.3 Results and Discussion

Figure 4 shows the average results for all subjects. The vertical axis shows the response time in seconds while the horizontal axis shows the stimulation points.

As Figure 4 shows, the response time was different for each stimulation point and about 1 second. Next, we calculated the deviation value for all response time in order to standardize the data. Table 1 shows the results for the deviation value of response time. When the response time was mean, the standardized

value was 50. If the value was larger than 50, it meant the response time was longer than the mean and vice versa.

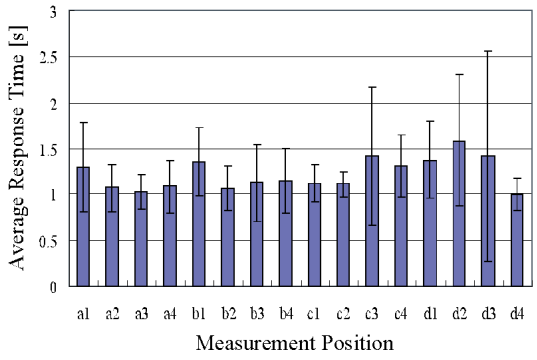


Figure 4: The response time.

Table 1: Standardized response time.

	1	2	3	4
a	47.9	38.2	34.7	40.7
b	53.7	38.6	41.1	40.3
c	44.4	45.6	45.2	53.3
d	53.5	54.0	42.9	34.8

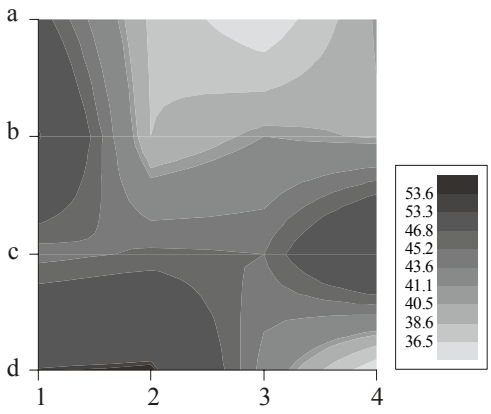


Figure 5: The standardized response time in contour graph.

Figure 5 shows the contour graph of standardized response time. From figure 5, it was found that the response time was short on position “a” and central area of palm. And it was also found that the response time was long on peripheral area of palm.

3 MEASUREMENT OF MECHANICAL IMPEDANCE

3.1 Experimental Procedure

The palmar mechanical impedance can be found from the power and the acceleration caused when the palm is vibrated at various frequencies. Therefore, we devised an experimental setup capable of inducing the vibration of the skin. We used a small vibrator, and measured the power and acceleration by means of an impedance head. The mechanical impedance of the skin was measured with the measurement setup shown in Figure 6. The vibrator outputted a sine wave vibration, and the vibration was relayed to the palm through the impedance head and the contactor pin. We used 15 measurement frequencies: 80, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 and 1000 Hz. A touch sensor was used to confirm that the palm touched the contactor pin. The same subjects in chapter 2 participated in this experiment. The subjects touched the pin with the palm of the right hand. When the pin touched the palm, the vibrator made the skin vibrate. The subjects were asked to control the pressure of their palm on the contactor pin by watching the output of the load cell. The power with which the skin pushed the pin was set to 50 gf. During the experiment, the subjects were asked to simply place their palm onto the contactor pin. The impedance of the palm was measured by the impedance head, and the data were inputted into a computer. The measurement was made as described in chapter 2 (Fig. 2).

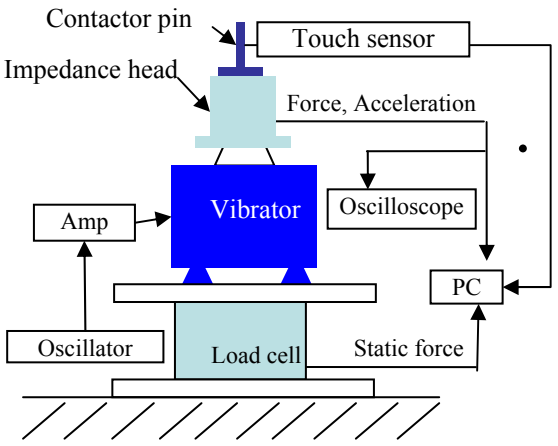


Figure 6: Mechanical impedance measurement device.

3.2 Results and Discussion

From the analyses of mechanical impedance, we divided them into two categories. The typical patterns of each category were shown in Figures 7 and 8. Figures 7 and 8 show the average results for all subjects at stimulation point a2 and d2, respectively. The vertical axis shows the impedance. The horizontal axis shows the frequency in Hz. The lower part of those graphs shows the imaginary part

of the impedance. The upper part shows the real part of the impedance.
Figure 7 shows that the value of imaginary part increased as the frequency increased. We called this pattern #1. On the other hand, the value of imaginary part increased and decreased as the frequency increased in figure 8. We called this pattern #2.

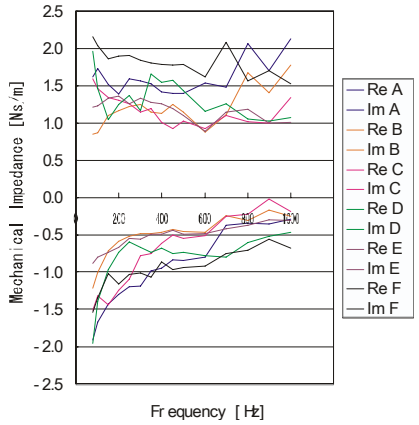


Figure 7: The impedance change (pattern #1).

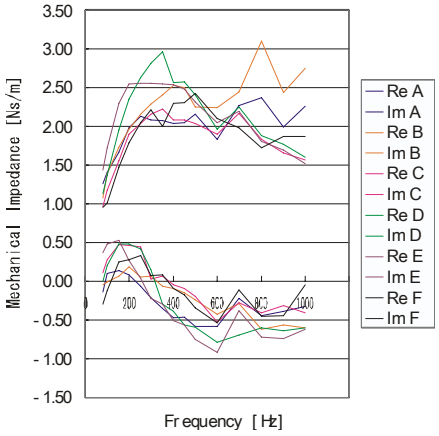


Figure 8: The impedance change (pattern #2).

Table 2 shows the categorized results for all stimulation points. One-asterisk shows that pattern #1 change was obtained in the stimulation point, while two-asterisks shows pattern #2. From table 2, it was found that the pattern #2 was obtained on peripheral palm where the hand was relatively thick and the pattern #1 was obtained on relatively thin part.

Table 2: Position dependence of impedance change.

	1	2	3	4
a	*	*	*	*
b	**	*	*	*
c	**	**	*	**
d	**	**	**	**

3.3 Comparison between Response Time and Mechanical Impedance

Figure 9 shows the combination between figure 5 and table 2. From figure 9, it was found the response time was relatively short at the area of pattern #1, while the response time was relatively long at the area of pattern #2. There seemed to be a relationship between response time and mechanical impedance. Therefore, we hypothesized that the response time could be obtained by using mechanical impedance.

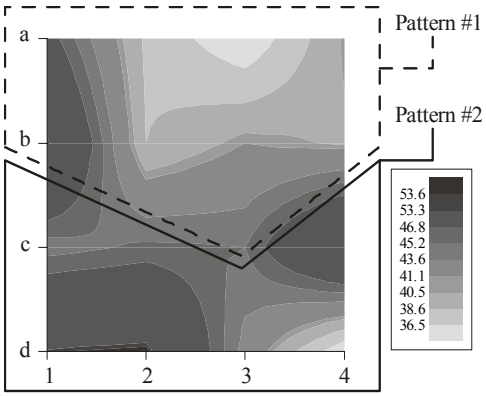


Figure 9: Comparison between response time and mechanical impedance.

4 CONCLUSION

We investigated the relationship between response time and mechanical impedance in this paper. In the near future, we are planning to make an energy conductive model by using mechanical characteristics in order to estimate thermal stimulation response.

REFERENCES

Horio, H., & Wada, C., 2005. Basic research into the development a Deaf-Blind communication device based on the use of thermal apparent motion, *SICE Annual conference 2005*, 348-352.

Itoh, K., Sakai, T., & Sakajiri, M., 2003. Assistive technology in the use of a PC for the deaf-blind(In Japanese), *WIT 2002 of the Institute of Electronics, Information and Communication Engineers*, 25-28.

Wada, C., & Wada, Y., 2003. A proposal for a communication device with a finger and a palm for the deaf-blind, *10th International Conference on Human - Computer Interaction*, 4, 281-285.

AUTHOR INDEX

AUTHOR INDEX (CONT.)

Proceedings of BIOSIGNALS 2008
International Conference on Bio-inspired Systems and Signal Processing
ISBN: 978-989-8111-18-0