34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Adaptive Affective Response Identification for Hearing Threshold Detection

Thomas E. Doyle[1], David Musson[2]

***Abstract*— Emotional arousal, or affective patterns, can be probed using observable bioelectric signals, in particular using the fluctuations of electroencephalographic potentials from the human scalp. Hearing impairment related to increased threshold of audio tone detection may cause the loss of intelligibility of speech resulting in an innate automatic emotional response. An adaptive support vector machine can be trained to identify a subject's unique affective response based upon an audiogram hearing test. This paper presents the efficacy of our model, initial SVM classification data, and discusses potential application.**

I. Introduction

The human body has numerous communication channels, transmitters and receivers, inter- and intra-communicating. The verbal inter-communication channel consists of speech, sound, and hearing and as with any communication system, it can be divided into a transmitter, a transmission medium, and a receiver. The physical and physiological process of human communication is well understood, but the mechanisms that control the speech articulators and hearing comprehension remain areas of active research.

II. The Communication Channel

A. Speech Production

Speech is composed of a collection of sequential complex sounds. The syntax and protocol of these sounds are governed by the rules of grammar that form symbols to convey information based on a cognitive lexicon. The formulation of these sounds is beyond the scope of this research; however, it has been proposed that we are born with the ability to rapidly acquire a meaning for innate concepts and that we use those meanings to develop a lexicon to communicate, using a universal grammar [1]. Rabiner states that the production of these sounds may be modeled with a time-varying linear system [2] that according to Parsons can be divided into two functions; excitation and modulation [3].

Production of speech requires energy. Speech energy is supplied from the expiratory phase of the breathing mechanism. Air flowing out of the lungs generates a steady flow of energy in one direction causing, initially, the vocal cords to oscillate, and then the air particles surrounding them. Fry best describes this as the vibration of a musician's lips on a brass wind instrument [4].

[1]T. E. Doyle is an Assistant Professor in the Faculty of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada `tdoyle at ieee.org`

[2]D. Musson is the Director of the Centre for Simulation-Based Learning and an Assistant Professor in Faculty of Health Science, McMaster University, Hamilton, Ontario, Canada `musson at mcmaster.ca`

The description of speech production becomes more complex because it can be described on three different levels:

1) Linguistic: study of the structure and nature of human speech (i.e. morphology, syntax, dialectology, phonology, etc.).
2) Acoustic: differences in the acoustic levels are due to dialect, physiological characteristics, and speaker mannerisms. These differences may be so great that it is not practical, or possible, to record the actual sounds. So instead, speech is characterized in terms of articulatory gestures. Alternatively, a spectrogram is used.
3) Articulatory: removes the physical characteristics, which cause the differences in acoustic levels, and allows the representation of speech to be done using a formal set of symbols (IPA - International Phonetic Alphabet).

A further confusion is the terminology relating these levels. At the linguistics level, a speech unit is called a *phoneme*, which is translated by an articulatory gesture into a *phone* at the acoustic level. The phoneme is the intended unit of language and the phone is the sound produced.

The propagation of speech sound waves is analogous to the sound produced when a tuning fork is struck. The vibration of the tuning fork produces a displacement of the immediate air particles forcing them into vibration. However the tuning fork will emit a single pure tone, or single frequency, while the human voice is composed of an infinitely possible combination of "tuning forks" producing complex tones, or mixture of frequencies.

There are two types of vibratory wave motion a particle can assume: transverse and longitudinal.

The transmission of sound through air is often compared to the ripple effect a pebble would produce on a still pond. If we consider the pebble's effect along a radial line, we see that the individual particle moves up and down, perpendicular to the outward motion of the wave; this is a transverse wave. This type of vibratory motion appears at first to resemble that of a tuning fork. However, this type of vibration can only occur on on a liquid's surface or within a solid and not in a gas, thus it cannot create sound waves (audible vibration) in air.

The vibration of the tuning fork causes compression and rarefaction of the contiguous air particles, causing them to oscillate along the line of travel; this is a longitudinal wave.

The compression and rarefaction motion of a single particle in a longitudinal wave causing the rise and fall of air

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

pressure will produce a sinusoidal waveform traveling omni-directionally from the source; expanding radially.

Any sound that reaches the ear drum is the result of compression and rarefaction of the air particles of a longitudinal waveform.

B. Reception and Perception

The information contained in human speech comes from more than the words that are used to convey our thoughts. Our auditory system must process speech by translating sound pressure waves into a sequence of electrical impulses that are passed through the nervous system into the brain. Different sounds in our language produce different frequencies of vibration in the ear and these vibrations contain information. The processing of this information is so automatic that we can not consciously be aware of the different vibrations and the information they contain.

The human ear consists of three main parts:

1) Outer ear: the visible part, the auditory canal and the eardrum,
2) Middle ear: the hammer, the anvil, and the stirrup bones, and
3) Inner ear: the cochlea, and auditory nerves.

Sound enters the auditory canal and applies pressure to the eardrum. The hammer, anvil and stirrup bones convey eardrum vibrations to the cochlea in the inner ear. These vibrations cause pressure waves to travel down the cochlea making the cochlea's tiny hair cells bend creating action potentials. The hair cells are attached to the auditory nerves and this information is transmitted to the brain.

III. Hearing impairment

Simply stated, hearing impairment can be described as a spectrum analyzer with a damaged channel [5]. Admittedly, this definition does not provide a sense of the many types and forms of hearing impairment, but it does provide insight to the resultant problem. The most common form of hearing loss is the mixed type of conductive and sensorineural damage. The degree of hearing impairment varies according to the division of the mix and by the method which the damage occurred.

Our focus is that of presbycusis-type impairment. Presbycusis is defined as the progressive increase of upper frequency threshold of hearing causing the threshold floor to rise causing reduced speech intelligibility due to the corruption of high frequency speech components. Similarly, if we consider the rise of the threshold floor at any frequency then we will experience similar speech intelligibility problems.

IV. Affect of Hearing Impairment

The type, method, severity, and age of onset of hearing loss has both psychological and social implications. The limited empirical literature finds hearing loss is associated with elevated rates of depression and anxiety [6]. Clinical study and experience suggest that because loss of hearing effects the unconscious and primitive level of hearing [7] and appeared to cause increased stress [8]; especially in adults with acquired hearing loss, in comparison to pre-lingually deaf. Also, the author's observations indicate a strong correlation between a hearing impaired individual's increased emotional stress and an instance of miscommunication; often this increased stress goes unnoticed by the impaired.

V. Affective Electroencephalography

Using scalp electrodes, past attempts to isolate brain current for singular cognitive functions has proven difficult, if not impossible. This is due to the diffuse nature of the brain's electrical signals through the organic medium of the scalp. The literature presents a generalized frequency range approach to classifying electrophysiological patterns, perhaps taking it's cue from the classification of the rhythms that have been well defined [9]–[12]: $\alpha, \beta, \gamma, \delta, and\ \theta$.

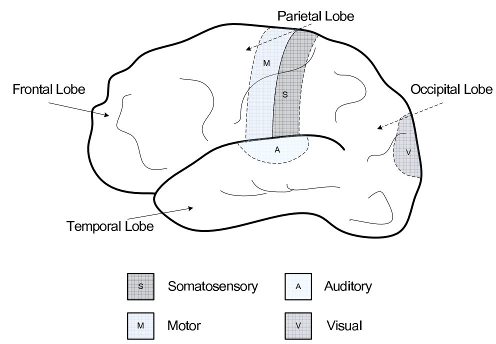

Fig. 1. Sensory associated regions of the human brain

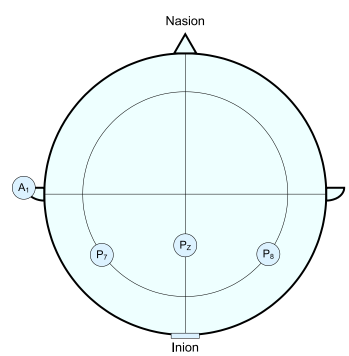

Fig. 2. An alternate electrode placement by Doyle *et al.*

Our electrode locations conform to the 10-20 system of positioning and we have expanded the horizontal range of the montage to reduce the influence by the occipital region (visual subsystem) and to increase the influence of the somatosensory (sensation) region of the brain. Regions of the brain are shown in figure 1. The standardized names for the chosen locations are: Inion, Pz, P7, and P8. Our reference is taken from the subject's left ear lobe.

VI. Methodology

There are well defined frequency bands that have been associated with cognitive and emotional attributes. Careful selection of features related to these attributes permits us to develop an adaptive model based upon the user's unique response.

Regardless of the attributes selected, the visualization and classification of such data becomes an increasing challenge proportional to dimensionality. Fundamental to the classification problem is the unknown distribution of the data, further complicated with the fact that real-world experimental data of this nature are often not linearly separable without significant misclassification error.

A. Electrophysiological Response to Auditory Tones: Threshold

When a hearing test is performed, an audiologist will measure the subject's hearing threshold. The threshold is determined by the subject's acknowledgement (or lack of) to series of short tones. These tones span the frequency range of human hearing and vary in amplitude. If a subject is deficient in any portion of their threshold response then they generally require a hearing assistive device.

An individual experiencing difficulty in communication due to a loss of audiological intelligibility exhibits an unconscious emotional response; an affective state.

The electroencephalogram is a measure of brain current. A synchrony of oscillating brain currents have been correlated with brain activity. Several of those correlated oscillations have well defined frequencies in brain research.

B. Support Vector Machine Classification

Often experimental data are inseparable using a linear classifier. This may be overcome by permitting some data to be misclassified, or by the use of a nonlinear classifier. A support vector machine (SVM) may be used to create nonlinear classification boundaries (hypersurfaces) by using a mapping (Φ) from input (attribute) space ($\mathbf{x}$) to feature space($\mathbf{z}$). The hypersurfaces are then created in feature space to partition our measured indicators. Measuring EEG data related to a user's unique response to threshold tone stimuli provides the training data for the support vector machine. Training time is minimal due to the innate response related to loss of intelligibility.

Figure 3 illustrates the process of performing the threshold experiments and how a single experimental data set was used to create training and testing data.

The implementation of the SVM requires the selection of several parameters: kernel function, associated kernel parameters, and a permissible margin of error. Each parameter was selected empirically. Our model used a Gaussian kernel because it is well suited to noisy experimental data [13]–[15] and also because of the nature of a Gaussian function relaxes the boundary restrictions of the SVM design.

The training data is extracted from the 8 kHz hearing threshold experiment in the first trail. This is illustrated in figure 3. The selected window of observation is two seconds, or 200 sampling points. To "*teach*" the SVM we require both input and output training data. However, we do not know the distribution of the output, but given the subject's acknowledgement we know that a response has just occurred. The subject is presented with several hearing threshold trials, of which we have labeled two such trails in figure 3. From the first trail the *training input* is extracted and for our *training output* we have partitioned it into fifteen possible segments. Training output partitions numbered 1 - 7 were used to investigate a response duration of 500 ms, and partitions 8-15 were investigated a refined response duration of 250 ms. The SVM is trained using all fifteen possible outputs. The training formulated the decision boundary (hypersurface) using a small number of training data points; support vectors. With the assertion that a response has occurred and that its occurrence is measurable, we evaluate a *testing set* of data.

As illustrated in figure **??** the testing set of data was the entire second trial which was independent of the first trial. Each subject acknowledgement in the testing data was extended back in time (P) to accommodate subject reaction latency. If our training data created a good general classifier (binary classifier or dichotomization) then we may expect that the testing data would have had very high percentage of correct classification.

Using the five electrode locations (Inion, Pz, P7, P8, lobe as reference) for raw EEG measurement, our parametric search examined all combinations of outputs data (1-15), subject reaction latency (P: 250 and 500 ms), upper limit on classification error (C: 1, 5, and 10), and the Gaussian kernel radial bias function (σ: 0.25, 0.50, and 1.00).

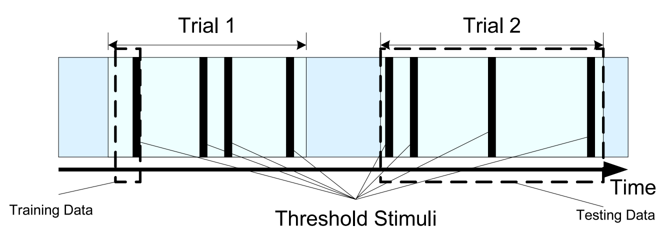

Fig. 3. Experimental trials and their relation to training and testing data

C. Subject Preparation

The montage required a total of five surface electrodes to be placed on the subject as presented in figure 2. The placement of the electrodes were done in accordance with the skin preparation and equipment decontamination procedures of the College of Physicians and Surgeons of Ontario [16] and the 10-20 system of electrode placement [11].

D. Experimental Setup

The number of subjects was three (n = 3), denoted as subject A, B, and C. The subjects were healthy males, without hearing impairment, between the ages of 25 to 35 years old.

The experiment required the subject to sit comfortably in a reclined position with eyes closed. All instruction and stimuli were presented though the headphones. Audio instruction explained the format of the experiment, method of acknowledgement, and duration of the experiment. The acknowledgement was button to begin, followed by ten seconds of silence. The subject is presented with 3 sets audio tones. Each audio tone set presented the series of 100 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, 5 kHz, 8 kHz, 10 kHz, 12.5 kHz, and 15 kHz. The order of presentation, tone duration and pauses were varied. Each tone is presented in discrete, but increasing steps of amplification. When a tone was detected, the subject was to acknowledge it via the acknowledgement button.

VII. Results

Correct SVM classification using raw EEG data from the five identified electrodes ranged from an average of 76.3% to 94.0%; however, false positive results were significant with the higher average classification. Duration of data partition (P: 250 and 500 ms) presented no improvement to classification. Increased computationally efficiency of classification ($\frac{\text{correct classification}}{\text{no. of support vectors}}$) occurred when upper limit on classification error (C: 1, 5, and 10) and the Gaussian kernel radial bias function (σ: 0.25, 0.50, and 1.00) were C = 5 and σ = 0.50, respectively.

VIII. Discussion

This study investigated the use of SVM machine learning in detection of hearing threshold using a subject's unique electroencephalographic emotional arousal, or affective response. Based upon sample size we do not have statistical significance; however, the range of successful classifications are encouraging for further investigation.

Electroencephalography for the measurement of affective patterns has inherent sources of error that shall be addressed in further work. Three identified areas for improvement: 1) reducing the noise EEG measurement, 2) remove muscular contraction artifacts, and 3) differentiate anticipation vs. actual response.

Recent research of electrocorticography (ECoG), a method of measuring electrical response where an electrode array is placed directly upon the subjects brain [17], [18], may provide significantly higher accuracy and selectivity. Example applications of ECoG are (a) implanted telemetry and control sensors for a prosthetic hand [19], (b) flexible electrode array overlaid directly on the human brain [20], and (c) Brain Computer Interface [21]. While ECoG is invasive, it much less invasive than using electrode arrays that are inserted into the brain to identify single unit neuron responses.

IX. Conclusion

The results from our model analysis are encouraging. By empirically determining a good set of parameters, our support vector machine generated a very high percentage of correct classification; however, false positives need addressed as noted. In reference to our model assumptions, we have validated that emotional responses (affective patterns) can be measured using bioelectric signals and that the loss of an innate attribute is sufficient to produce a reliable and measurable response without explicit operant conditioning. While this detection method is well suited to determine a hearing threshold for the improved fitting of custom hearing prostheses, it is expected that this methodology will find best application embedded in the digital hearing aid for specific acoustic environment response and adaptation.

This dual feedback mechanism (subject and embedded computation adapting to each other) with the subject-in-the-loop offers an exciting application for machine learning and human rehabilitation.

References

[1] N. Chomsky, *Language and Problems of Knowledge*, ser. The Managua Lectures. The MIT Press, 1988.

[2] L. R. Rabiner, *Digital Processing of Speech Signals*. Prentice-Hall, 1978.

[3] T. Parsons, *Voice and Speech Processing*. McGrawHill, 1986.

[4] D. B. Fry, *The Physics of Speech*, ser. Cambridge Textbooks in Linguistics. Cambridge University Press, 1979.

[5] A. M. Engebretson, "Benefits of Digital Hearing Aids," *IEEE Engineering in Medicine and Biology*, pp. 238–248, April/May 1994.

[6] B. G. Frankel, "Adult-onset hearing impairment: Social and psychological correlates of adjustment," Ph.D. dissertation, Department of Epidemiology and Biostatistics, Faculty of Science, University of Western Ontario, London, Ontario, 1981.

[7] D. A. Ramsdell, *Hearing and Deafness*, 4th ed. Holt, Rinehart and Winston, 1978.

[8] H. Myklebust, *The Psychology of Deafness*, 2nd ed. Grune and Straton, 1964.

[9] K. A. Kooi, R. P. Tucker, and R. E. Marshall, *Fundamentals of Electroencephalography*, 2nd ed. Harper and Row, 1978.

[10] K. Sauve, "Gamma-band synchronous oscillations: Recent evidence regarding their functional significance," *Consciousness and Cognition*, vol. 8, pp. 213–224, 1999.

[11] B. J. Fisch, *EEG Primer: Basic Principles of Digital and Analog EEG*, third revised and enlarged ed. Elsevier, 1999.

[12] C. H. Vanderwolf, "Are neocortical gamma waves related to conciousness?" *Brain Research*, vol. 885, pp. 217–224, 2000.

[13] A. I. Belousov, S. A. Verzakov, and J. vonFrese, "A flexible classification approach with optimal generalisation performance: support vector machines," *Chemometrics and Intelligent Laboratory Systems*, vol. 64, pp. 15 – 25, 2002.

[14] K. Müller, S. Mika, G. Rätsch, K. Tsunda, and B. Schölkopf, "An introduction to kernel-based leraning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181 – 202, March 2001.

[15] V. Kecman, "Support vector machine basics," The University of Auckland, School of Engineering, Tech. Rep., April 2004.

[16] D. J. Klass, "Guidelines for clinical practice & facility standards: Electroencephalography," College of Physicians and Surgeons of Ontario, Tech. Rep., September 2000.

[17] A. E. Schultz and T. A. Kuiken, "Neural Interfaces for Control of Upper Limb Prostheses: The State of the Art and Future Possibilities," *PM & R : J. Injury, Function, and Rehabilitation*, vol. 3, no. 1, pp. 55–67, 2011.

[18] P. Brunner, A. L. Ritaccio, J. F. Emrich, H. Bischof, and G. Schalk, "Rapid Communication with a P300 Matrix Speller Using Electrocorticographic Signals (ECoG)," *Frontiers in Neuroscience*, vol. 5, no. 0, p. pp.9, 2011.

[19] R. Weir, P. R. Troyk, G. A. DeMichele, D. A. Kerns, J. F. Schorsch, and H. Maas, "Implantable Myoelectric Sensors (IMESs) for Intramuscular Electromyogram Recording," *IEEE Trans. Biomed. Eng.*, vol. 56, pp. 159–171, 2009.

[20] R. Oostenveld, "ECoG Brain Dynamics in High-Resolution Recordings," 5th BCI2000 Workshop & International Workshop on Advances in Electrocorticography, Bolton Landing, NY, USA, Tech. Rep., October 2009.

[21] E. C. Leuthardt, "Utilization of BCI for Potential Treatment of Hemispheric Stroke," 5th BCI2000 Workshop & International Workshop on Advances in Electrocorticography, Bolton Landing, NY, USA, Tech. Rep., October 2009.