35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Exploratory Study of EEG Burst Characteristics in Preterm Infants

Zhayida Simayijiang*, Sofia Backman†, Johannes Ulén*, Sverre Wikström‡, Kalle Åström*

*Centre for Mathematical Sciences, Lund University, Sweden
{zhayida,ulen,kalle}@maths.lth.se

† Department of Clinical Neurophysiology, Skåne University Hospital, Sweden
sofia.backman@skane.se

‡ Center for Clinical Research, County Council of Värmland, Sweden
sverre.wikstrom@liv.se

***Abstract*— In this paper, we study machine learning techniques and features of electroencephalography activity bursts for predicting outcome in extremely preterm infants. It was previously shown that the distribution of interburst interval durations predicts clinical outcome, but in previous work the information within the bursts has been neglected. In this paper, we perform exploratory analysis of feature extraction of burst characteristics and use machine learning techniques to show that such features could be used for outcome prediction. The results are promising, but further verification in larger datasets is needed to obtain conclusive results.**

I. Introduction

Preterm birth is an important cause of neonatal morbidity and mortality world wide [8]. Increasing numbers of extremely preterm born infants are surviving, however with high incidence of neurodevelopmental impairment [9]. There is an urgent need for improved monitoring of brain function in this population in order to identify brain damage, to direct care and in a future potentially also to provide early neuroprotective interventions. Electroencephalography (EEG) is a recording of brain derived voltage gradients over the scalp and provides an approach to assessment of brain function in real time.

The normal preterm EEG is characterized by high voltage activity bursts, also named spontaneous activity transients (SATs), alternating with relative inactivity (interburst intervals, IBI) of low voltage [7], see Fig. 1. Such bursts are crucial for early brain development and may provide an opportunity for studies of early brain function development [10]. It should be empathized that bursts (or SATs) in this context refer to a normal physiological event, distinct from the pathological burst-suppression pattern of full term infants or older subjects after major cerebral insults [3].

There is so far no clear definition or even description of preterm bursts. In previous literature, bursts have often been numerically defined as activity of a certain duration and amplitude above a defined threshold. This is in contrast to routine clinical detection of bursts which relies on visual pattern recognition. Consequently, detection of bursts may vary somewhat among raters [4], two different labelings of the same signal are possible and this can influence the interburst interval, shown in Fig. 1.

Previous studies have shown that reduced numbers of bursts and prolonged IBIs during the first postnatal days are indicators of acute brain dysfunction predictive of poor neurological and cognitive outcome [5], [11]. In future clinical studies, we aim to test the hypothesis that not only the durations (or number) of bursts and IBIs contain clinically valuable information, but also the detailed characteristics of bursts themselves may contain such information as well. In this paper, we study a number of burst features and use machine learning to examine the predictive properties of features in a pilot sample. A study flow chart is shown in Fig. 2.

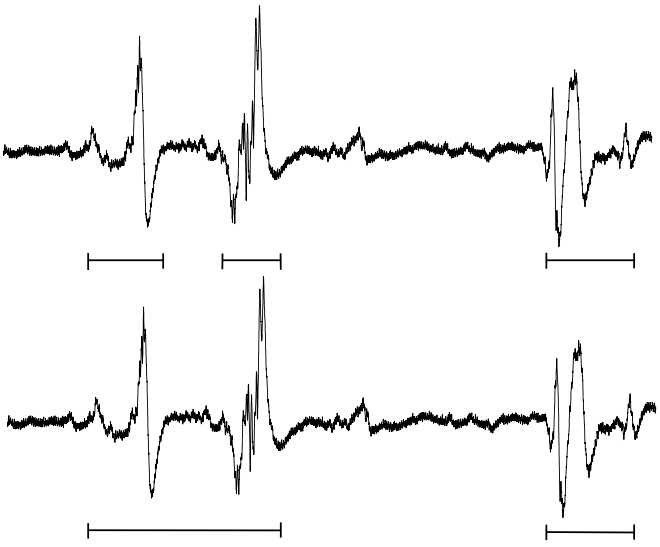

Fig. 1. Two possible labelings of burst and interburst intervals in the same recoding epoch. Bursts are underlined.

II. Data and preprocessing

The dataset consisted of one-channel EEG recordings during the first three postnatal days of 14 previously described extremely preterm infants belonging to a larger study cohort [11], [2], 23-30 weeks gestational age. Eight infants had good outcome and six had poor outcome, defined as neurodevelopmental impairment according to psychological testing and neurological examination at two years age. Written informed consent was obtained from all parents and the Regional Ethics Committees in Lund approved the research protocol. Voltage gradients between standardized electrode locations

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

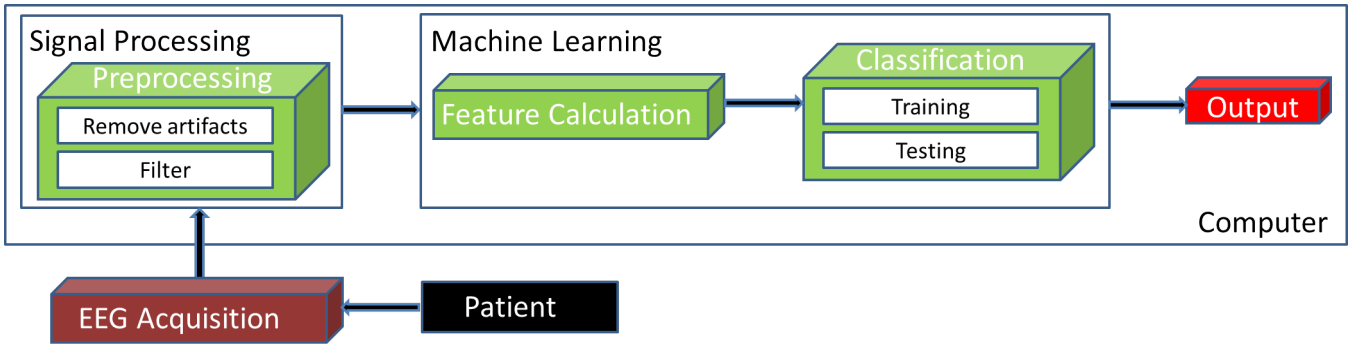

Fig. 2. A flow chart of the study.

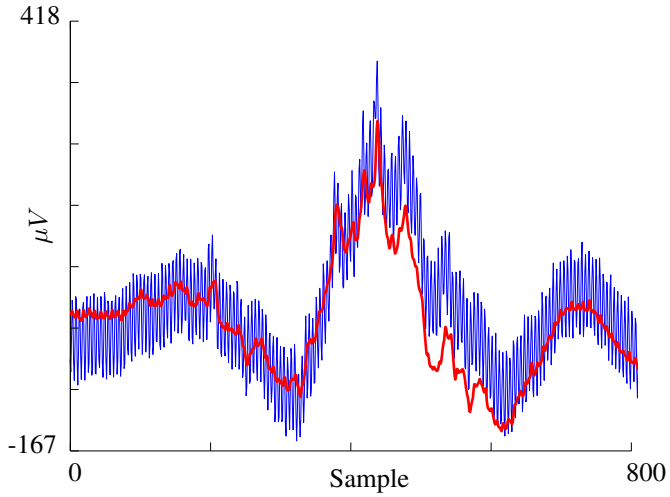

Fig. 3. Signal before and after filtering. The thin blue line is the original signal. The thick red line is the signal after filtering. Note that the low frequency components have also been removed from the signal by the band-pass filter.

P3 and P4 (reference Cz) according to the International 10-20 System were recorded, at sampling frequency 256 Hz using a Nervus/NicOne 3.3 EEG system with U16 amplifier (Natus Medical Incorporated, San Carlos, CA). Recordings were reviewed (filter between 0.5 and 70 Hz, sensitivity 100 microvolt/cm, timebase 30 mm/s) by two independent experienced clinical neurophysiologists who marked at least 100 bursts from each patient during the second postnatal day. Artifacts (e.g., technical interference or movement artifacts) were also marked and removed during preprocessing of the signal. To remove noise, we applied a notch filter at 50 Hz, together with a band pass filter on the interval from 0.5 to 70 Hz. Fig. 3 shows the signal before and after filtering.

III. Features

In this article, six features are calculated trying to capture different aspects of the signal. For each patient each burst is handled individually, and calculated these six features for each burst. Basically, we choose these features in the sense of they complement each other, and independent from one another. Two of these features (Spectral edge frequency and Shannon entropy) have been previously described by Löfhede [6], for the purpose of identification and segmentation of pathological burst-suppression patterns. We also introduced four novel features as presented below.

A. Sharpness

There are quick turns in burst signals compared to interburst. Thus for each burst signal, we find local minima and maxima of the signal, and calculate the angle of each peak, then use the mean value of these as a feature. This is shown in Fig. 4. Since interburst is more flat than burst, we believe that sharpness is a characteristic feature.

B. Number of peaks

For each burst, we count the number of local extrema of the signal. As aforementioned, burst signals have more peaks as compared with interburst.

C. Spectral edge frequency (SEF)[6]

SEF 95 is the frequency under which 95 % of the signal power resides, it is calculated based on Fourier transform.

D. Ripple Sharpness

The bursts are typically built up by very slow (low frequency) waves with superimposed high frequency components [10], here called ripples. To capture this phenomenon the signal is decomposed into two components. The slow wave is approximated by applying a median filter of length 15. The ripples are then extracted as the difference between the median filtered signal and the original signal, see Fig. 5. The feature we used is the mean angle of the peaks of the ripples.

In future analysis, we will compare ripples during bursts with frequencies of the interburst intervals, to ascertain the biological signal and to be able to filter out even more noise.

E. Burst duration

The elapsed time for each burst.

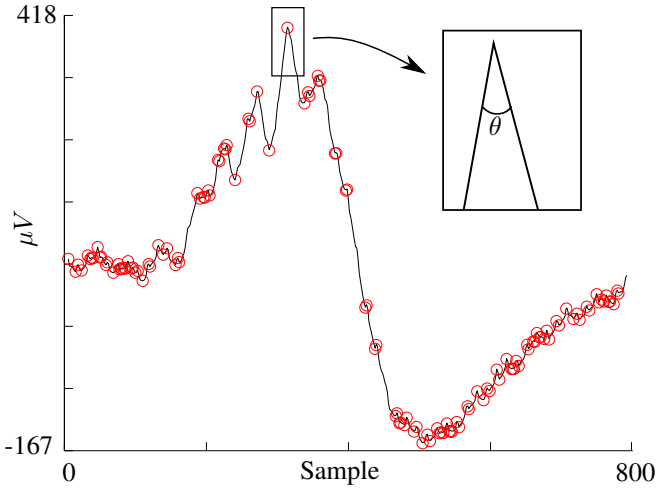

Fig. 4. The peaks and angles of a burst. Here EEG signal is shown in black and each red circle indicates a local extremum. The angle, θ is emphasized.

F. Shannon entropy [6]

Entropy is a measure of uncertainty associated with a random variable and used as a measure of signal complexity. For measure Shannon entropy of signal S, we take the signal into $n = 20$ bins and calculate

$$-\sum_{i=1}^{n} p(S_i) \log p(S_i), \tag{1}$$

where $p(S_i)$ is the probability mass function of S_i.

IV. Classification

In this paper, we use machine learning techniques for automatic classification of outcomes based on features and for assessment of feature usefulness.

The usual machine learning pipeline consists of calculating feature vectors based on the input data. Ideally each feature individually depends only on corresponding class. In practice the correlation between features and classes are complex and this is where classifiers comes in.

The classifier uses the features as input and makes a prediction of which class the data comes from. In this study we have opted to use random forests as our classifier.

Random forest [1] has several advantages. It is fast and relatively robust to noise. It is also possible to assess how important the different features are and how certain the classifier is on each decision.

Random forest belongs to the class of bagging classifiers which combines the results from many simple classifiers into one classifier. As the name hints random forest bags decision trees. Each decision tree is constructed by randomly sampling two thirds of the data. The last one third of the data is left out as the so called out-of-bag (OOB) data. The final classifier is simply the majority vote of the bagged decision trees.

Each tree gives a vote on which class it believes the data belongs to and by counting the votes a measure of confidence can be constructed.

The importance of any features is measured by scrambling the feature, and then calculating how much worse the prediction becomes (on the OOB set). In Section V, the importance is measured as mean decrease in accuracy (on the OOB set).

Assume that we have a two class problem with equal amount of examples from both classes and one perfect feature. The perfect feature would give a 100 % correct classifications, and scrambling the feature would lowering the importance, thus would be give an accuracy of 50 % on average (random assignment).

Due to the small size of the dataset, we use the leave-one-out paradigm for training and testing. One patient was held out and used for testing while all other patients were used to build the classifier. For each burst, we keep the votes of every tree. In the end, we let the majority vote over all tree to make the classification. This was repeated for each patient, thus producing 14 classifiers.

V. Results

In Table I, we list the features used for classification and their importance in descending order, the accuracy was 71.4 % (i.e., correctly classified ten out of fourteen patients). The importance is calculated as mean over each of the 14 classifiers constructed during leave-one-out testing. We can see promising decrease in classification error for some features, while other features like Shannon entropy probably have no positive effect on the classification. Note that in this test we use no information from the interburst length characteristics.

TABLE I
Classifying using our features. Importance is measured as mean decrease in accuracy.

Features	Importance (accuracy)
Sharpness	14 %
Spectral edge	7.0 %
Ripple Sharpness	6.6 %
Number of peaks	3.9 %
Burst duration	3.2 %
Shannon entropy	1.0 %

As a reference we also train a random forest only on interburst interval information. Two aspects of the interburst interval in relation to outcome in extremely preterm infants have previously been examined: The median duration of interburst intervals and the fraction of the total recording duration consisting of interburst intervals [11]. Using only these two as features for the random forest and leave-one-out classification we obtain an accuracy of 57.1 % (i.e., correctly classified eight out of fourteen patients), the importance is given in Table II.

VI. Conclusion and Future work

In this paper, we have studied EEG signal burst characteristics in a pilot dataset from 14 extremely prematurely born infants. From each burst six features were extracted and random forest techniques were used for classifying

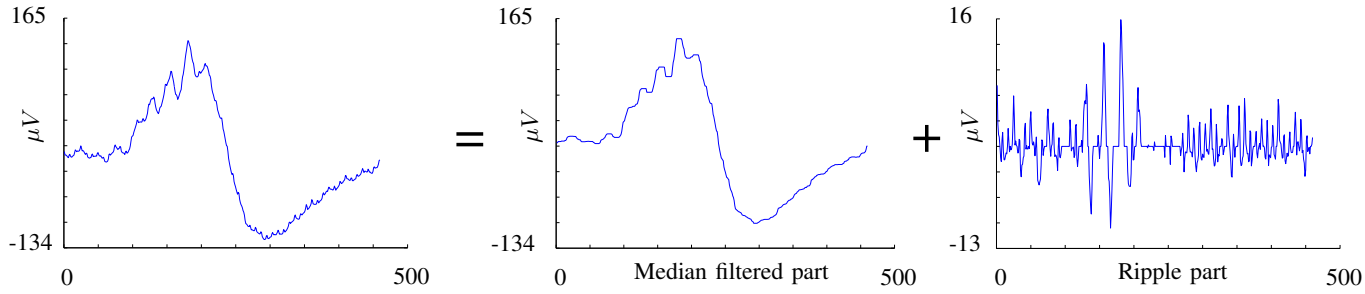

Fig. 5. Example of what the ripple looks like. The signal is decomposed into two parts and features are calculated on the ripple part.

TABLE II
CLASSIFYING USING FEATURES OF INTERBURST INTERVALS. IMPORTANCE IS MEASURED AS MEAN DECREASE IN ACCURACY.

Features	Importance (accuracy)
Median duration	4.7 %
Fraction	1.9 %

outcome based on the features of each individual burst, and for all bursts of each patient. The results are promising with precision of about 70 % on leave-one-out tests, which compares favorably to only considering durations of interburst intervals. This is a preliminary indication that the physiological EEG bursts do contain prognostic information. The study sample is, however, too small to draw conclusions about the exact predictive ability of this novel method.

If combined with development of automated burst detection [6], [4], the present method could make it possible to follow development of burst characteristics over time, i.e., real time monitoring. Such technical support would facilitate an objective analysis sensitive to changes that today may be overlooked in visual pattern recognition assessment. It may also support earlier detection of neural dysfunction when there still is time to act clinically. Consequently, further development of reliable burst detection methods, feature extraction and machine learning methods for burst and interburst analysis may improve outcome prediction and diagnostic ability.

Thus for the future we would like to (i) develop fast, automatic and reliable burst detection methods, (ii) collect a larger dataset of patient data with known outcomes and (iii) further develop and evaluate feature extraction methods and machine learning methods for the analysis of both burst characteristics and interburst interval distribution analysis for improved outcome prediction and diagnostic assistance.

VII. ACKNOWLEDGMENTS

The authors would like to thank prof. Lena Hellström-Westas and prof. David Ley for providing EEG data, and prof. Ingmar Rosén and associate prof. Gert Andersson for valuable input regarding EEG physiology and manual EEG burst marking, respectively and prof. Fredrik Kahl for valuable input on modeling and machine learning.

Zhayida Simayijiang was supported by the strategic research programme eSSENCE, and by the European Research Council (GlobalVision grant no. 209480). Johannes Ulén was supported by Swedish Foundation for Strategic Research (SSF) through the programme Future Research Leaders, and by the European Research Council (GlobalVision grant no. 209480). Sofia Backman was supported by Skåne county council's research and development foundation, Stiftelsen Samariten and the Linnéa and Josef Carlsson's foundation. Sverre Wikström was supported by grants from the County Council of Värmland, Sweden.

REFERENCES

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] I. Hansen-Pupp, H. Hövel, A. Hellström, L. Hellström-Westas, C. Löfqvist, EM. Larsson, F. Lazeyras, V. Fellman, PS. Hüppi, D. Ley. Postnatal decrease in circulating insulin-like growth factor-I and low brain volumes in very preterm infants. *J Clin Endocrinol Metab, 96(4):1129-35. doi: 10.1210/jc.2010-2440. Epub 2011 Feb 2*, 2011 Apr.

[3] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. Classification of burst and suppression in the neonatal electroencephalogram. *Neural Engineering*, 2008.

[4] K. Palmu, S. Wikström, E. Hippeläinen, G. Boylan, L. Hellström-Westas , S. Vanhatalo. Detection of 'EEG bursts' in the early preterm EEG: visual vs. automated detection. *Clin Neurophysiol, 121(7):1015-22. doi: 10.1016/j.clinph.2010.02.010. Epub 2010 Apr 14*, 2010 Jul.

[5] L. Hellstrom-Westas , H. Klette , K. Thorngren-Jerneck, I. Rosen. Early prediction of outcome with aEEG in preterm infants with large intraventricular hemorrhages. *Neuropediatrics, 32(6):319-324*, Dec 2001.

[6] J. Löfhede. *The EEG of the neonatal brain: classification of background activity*. PhD thesis, Göteborg: Chalmers University of Technology, 2009.

[7] M. Andre, MD. Lamblin, AM. d'Allest et al. Electroencephalography in premature and full-term infants. Developmental features and glossary. *Neurophysiol Clin, 40(2):59-124*, May 2010.

[8] S. Beck, D. Wojdyla, L. Say et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bull World Health Organ , 88(1):31-38*, Jan 2010.

[9] S. Johnson, J. Fawke, E. Hennessy, et al. Neurodevelopmental disability through 11 years of age in children born before 26 weeks of gestation. *Pediatrics, 124(2):e249-257*, Aug 2009.

[10] S. Vanhatalo, K. Kaila. Development of neonatal EEG activity: from phenomenology to physiology. *Semin Fetal Neonatal Med, 11(6):471-478*, Dec 2006.

[11] S. Wikström, IH. Pupp, I. Rosén, E. Norman, V. Fellman, D. Ley, L. Hellström-Westas . Early single-channel aEEG/EEG predicts outcome in very preterm infants. *Acta Paediatr, 101(7):719-26*, 2012 Jul.