Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

DETECTION OF BURSTS IN THE EEG OF POST ASPHYCTIC NEWBORNS

J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Abstract— **Eight features inherent in the electroencephalogram (EEG) have been extracted and evaluated with respect to their ability to distinguish bursts from suppression in burst-suppression EEG. The study is based on EEG from six full term infants who had suffered from lack of oxygen during birth. The features were used as input in a neural network, which was trained on reference data segmented by an experienced electroencephalographer. The performance was then evaluated on validation data for each feature separately and in combinations. The results show that there are significant variations in the type of activity found in burst-suppression EEG from different subjects, and that while one or a few features seem to be sufficient for most patients in this group, some cases require specific combinations of features for good detection to be possible.**

I. INTRODUCTION

The burst-suppression (BS) pattern is one of several indicators of severe pathology in the electroencephalogram (EEG) signal that may occur after brain damage, caused by e.g. asphyxia [1, 2]. Certain characteristics of this pattern can provide the clinicians with important information about the recovery of the patient and it is thus important in the adjustment of the treatment. Examples of important characteristics of the BS pattern are the length of the burst and suppression intervals, and the spectral content of the bursts [3, 4].

In practice, most EEG recordings are evaluated by experienced neurophysiologists through visual inspection of the unprocessed signal [5]. Some attempts towards automatic detection have been made, most of them targeting burst-suppression induced by anesthesia [6, 7].

Our goal is to develop tools that can be used for automatic classification of BS caused by various causes, for example perinatal asphyxia (insufficient oxygen/dioxide ventilation around the time of birth). The target patient group is newborns, and a possible future application is a monitoring system for neonates in intensive care.

This paper is focused on evaluating a number of features inherent in the EEG signal, with respect to their ability to distinguish between burst and suppression activity in BS EEG. These features are then used as inputs to a neural network for the actual classification.

II. METHODS

This study was performed on EEG from six full term infants having suffered from perinatal asphyxia. Each subject contributed with a continuous recording of 6-40 minutes, selected and classified by an experienced electroencephalographer (MT). The length was chosen to include at least 10 bursts. Eight channels were used, and the data was digitized at a sampling rate of 200 Hz. The mean of the EEG signal was subtracted, the signal was band pass filtered 0.5 to 20 Hz, and notch filtered at 50 Hz to reduce power line interference, before feature generation. The reference data had a time resolution of one second.

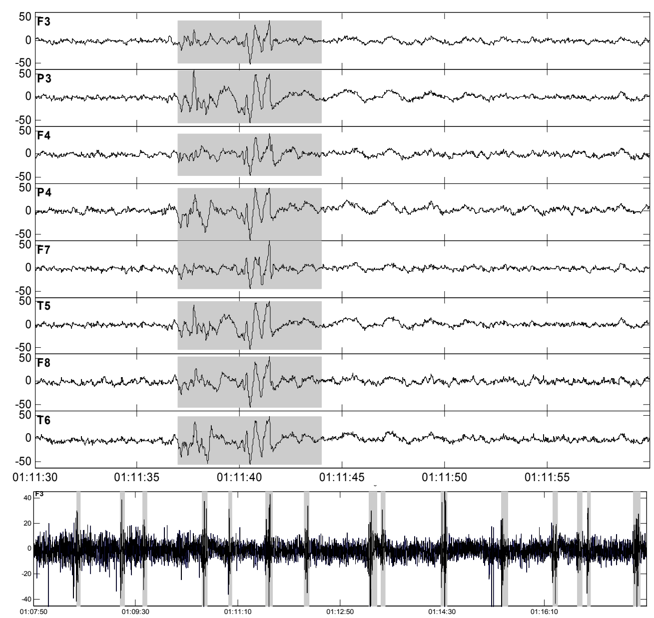

Fig. 1. The top plot is an example of 30 seconds of burst-suppression EEG. The bottom plot shows ten minutes of the same signal, but only one channel. The shaded areas are bursts identified by an electroencephalographer.

A selection of the recorded burst-suppression EEG is displayed in Fig. 1. Eight features, most of which have been used for detection of BS in adult subjects under anesthesia [7, 8], were selected (Table I). The features were calculated for overlapping segments of the raw signal. The segment

Manuscript received April 3, 2006. This study was supported by a grant from the BIOPATTERN EU Network of Excellence, EU Contract 508803, and the Margarethahemmet Foundation.

J. Löfhede is with the School of Engineering, University College of Borås, Borås, Sweden. (phone: +46(0)31-772 3713; fax +46(0)31-772 1782; e-mail: johan.lofhede@hb.se.)

N. Löfgren and K. Lindecrantz are also with School of Engineering, University College of Borås.

A. Flisberg and I. Kjellmer are with the Department of Pediatrics, Göteborg University, The Queen Silvia Children's Hospital, Göteborg, Sweden.

M. Thordstein is with the Institute of Neuroscience and Physiology, Section for Neuroscience and Rehabilitation, Sahlgrenska University Hospital, Göteborg, Sweden.

length was one second, motivated by the fact that bursts were considered to be at least one second long. The overlap was 0.75 s, producing an output feature signal with an effective sampling rate of 4 samples per second.

TABLE I
FEATURES SELECTED FOR EVALUATION OF THEIR ABILITY TO DISTINGUISH BURSTS FROM SUPPRESSION IN BURST-SUPPRESSION EEG.

Feature	Description
Spectral Edge Frequency (SEF95) [6]	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on rectangular windows of the signal
3 Hz power	Power in a one Hz wide band centered at three Hz
Median [9]	Median absolute value
Variance ($s^2(x)$) [9]	$s^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x-\mu)^2$ Where x is a time series, μ is the sample mean of x
Skewness ($E^3(x)$) [9]	$E^3(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\mu)^3}{\sigma^3}$ σ is the standard deviation of x. x and μ as above
Kurtosis ($E^4(x)$) [9]	$E^4(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\mu)^4}{\sigma^4}$ x, μ and σ as above
Zero crossings [6]	Rate of zero crossings
Shannon entropy (H_{Sh}) [10]	$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$ $p(I_1) \ldots p(I_U)$, p_U is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between the maximum and minimum value of the signal in the window. H_{Sh} is a measure of uncertainty of a random variable.

After feature generation, the feature signals were combined per feature by taking the median over the eight channels. This was motivated by the fact that BS is considered to be a global phenomenon appearing in all EEG channels, and using the median over the channels removes disturbances present on one or a few of the channels. This also reduces the computational complexity as compared to feeding the network with all channels.

The feature signals were used as inputs to a feed-forward neural network with error back-propagation training [11], which was implemented using the neural network toolbox in Matlab. The training of the network was performed using the segmented reference data. The feature signals were fed to the network one at a time, and then in combinations. The number of input nodes was set by the number of features used in each experiment. Ten hidden nodes were used, chosen by empirical testing, and one output node. Ten instances of the network were trained for each feature, and the best one was chosen for the evaluation. This technique was used to reduce the risk of using a network that failed to converge when trained. The training of neural networks uses a randomly initiated starting network, which then is trained using randomly selected samples from the training set. The training algorithms move in a complex landscape, trying to find a minimum error state. However, this landscape usually contains multiple local minima, and to reduce the risk of "getting stuck" in a local minimum, a number of separate instances of the network are trained, and then the instance with the lowest classification error is used.

In order to maximize the use of the limited number of patients, the data was not divided into fixed training- and testing sets. Instead, leave-one-out cross-validation was used, meaning that the data from one patient was used for testing, and the other five were used for training. This process was then repeated for all patients. In order to give each patient an equal chance to influence the network training, the shorter records were repeated so that they got the same length as the longest one. This equalizes the amount of data from each patient, but it does not take into account the number of bursts provided by the patients.

An output sample from a neural network depends only on one input feature vector, and does not take previous or subsequent samples into consideration. The detection signal is therefore noisy, with a lot of short detections. Using knowledge of the typical length of a burst, an iterative smoothing algorithm was designed that converts groups of short detections into longer continuous detections, while short isolated detections are removed. This was found to increase the performance considerably.

For performance evaluation, sample wise sensitivity and specificity were used as measures. Sensitivity was defined as the percentage of the burst samples that were correctly classified as bursts, and specificity as the percentage of the remaining samples that were not classified as bursts. As a measure of performance for choosing which networks to keep when training multiple instances, the sum of the sensitivity and specificity was used.

The features were ordered with the best single feature as a starting point. Initial experimental results showed that most patients were not affected very much by adding features, with the exception of patient six. Therefore, features were added in an order that increased the performance for patient six, while trying not to degrade the performance for the other ones.

III. RESULTS

In Fig. 2 the eight feature signals are compared. The mean and standard deviation of the feature signals have been calculated for the burst and suppression parts separately. The feature signals have been normalized by subtracting the mean and dividing by the standard deviation to simplify comparison. The figure shows that although the estimated means for burst and suppression of most feature signals differ, their standard deviations overlap considerably.

The results in Fig. 3 show that there are large differences between the patients with regard to how well the features work in terms of sensitivity. Using these results, the best feature was used as a starting point for the feature order for the test shown in Fig. 4. Here, both individual and total sensitivity and specificity are shown for increasing numbers of features.

Fig. 5 demonstrates the difference between sample wise sensitivity and burst wise sensitivity. Although the sample wise sensitivity in this case is merely 50%, all but two bursts have been detected.

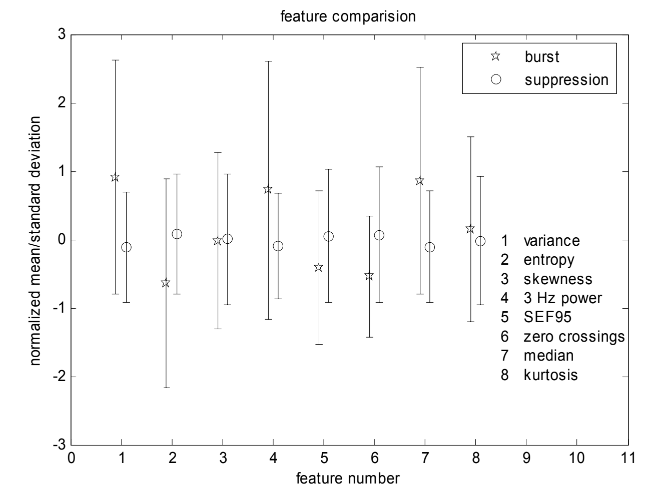

Fig. 2. The triangles and stars represent the mean of the feature signals for suppression and burst periods respectively. The lines show the standard deviations around the means. Data from all six patients are included.

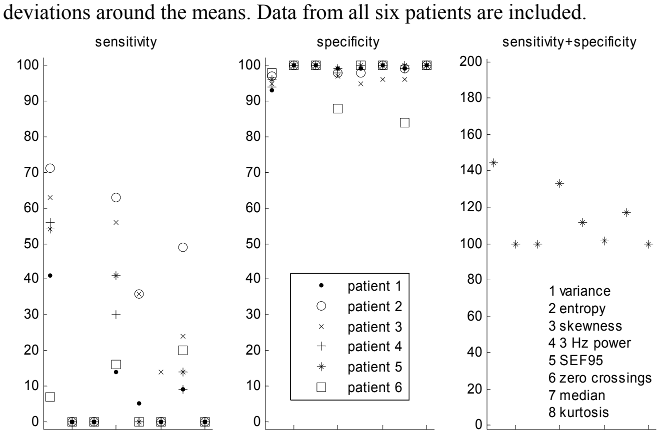

Fig. 3. Sensitivity and specificity achieved when the feature signals have been run through a neural network, one at a time. In each case, five networks have been trained, and the best one has been chosen. The sensitivity is in many cases very low, but this is partly due to the way the sensitivity is calculated. See Fig. 5 for comparison.

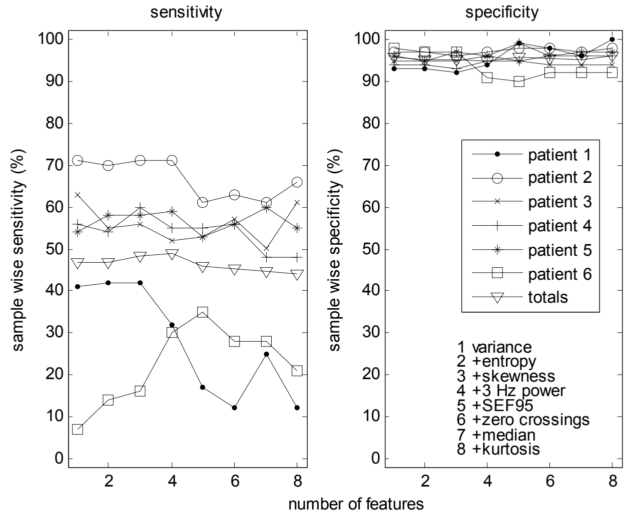

Fig. 4. Sensitivity and specificity after the feature signals were run through a neural network. The curves show the performance for the six patients for increasing numbers of features. The curve with the triangles shows the total performance for the six patients. In each case, ten networks were trained, and the best one with regard to the training set was chosen. The first point along the x-axis shows the result when feeding the network with the best feature, the variance. The subsequent points show the results when more features are added one by one until all eight features are included.

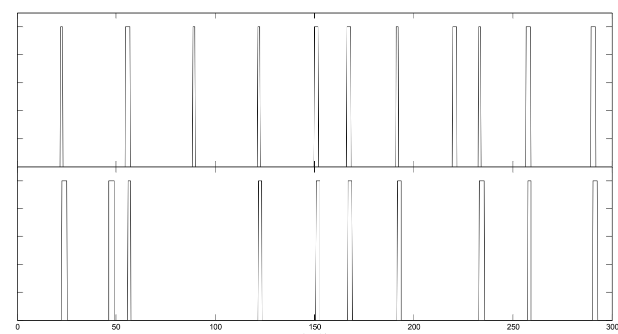

Fig. 5. Example of detection using four features on patient 5. The top plot shows the reference data, the plot below shows the detections. It can be seen that 9 out of 11 bursts in the reference data have been detected and that there was one false detection.

TABLE II
PATIENT BS RATIO.

Patient Number	Burst Percentage
1	24 %
2	13 %
3	6 %
4	27 %
5	26 %
6	4 %

Table II shows the burst contents in the signals from the different subjects. Note that the data from patient six, who has the worst detection performance, has a significantly lower burst content that most of the others.

IV. Discussion

In all calculations of this study, all eight available channels have been used. However, since bursts are mainly global phenomena, it should in principle be possible to detect them using only one channel. The drawback of using fewer channels is that using less data makes the system more sensitive to disturbances affecting only some of the channels, like bad electrode-skin coupling and artifacts. As an alternative to using the median of the input feature signal, one network could be trained for each channel, and then the final decision could be made by counting the number of detections from the different channels. This would however increase the computational complexity for the network training.

The method to calculate sensitivity and specificity used here is rather strict; every missed sample lowers the measured performance. The one-second time resolution used in the reference data probably contribute to a lowering of the calculated performance. Alternatively, the number of bursts detected divided by the total number of bursts could be used to define sensitivity, which would lead to higher scores. However, this would not at all take into consideration the length of the bursts, or the accuracy of the detection of the start- and endpoints.

The specificity for many of the features is very low (Fig. 3), but they still contribute to a higher sensitivity for patient six when used in combination with the variance (Fig. 4). However, adding the features in descending sensitivity order does not necessarily improve the result. An explanation for this may be that the best features share a lot of common information, and that combining them does not add useful information for the burst detection.

For most cases, the specificity is higher than the sensitivity. This is probably due to the fact that the burst periods are generally much shorter than the suppression periods, which means that the network has more examples of suppression to learn from than it has examples of bursts. The detections are often short, meaning that even a large number of false detections contribute little to a worsening of the specificity.

Given the small amount of data available, not much can be said about the actual performance of the neural networks. For most patients in the data set the detection works well when using only one feature, but for patient six the performance is very low when using one feature, and increases drastically when adding more features. At the same time, the performance for patient one decreases, while all other remain almost constant, even though the EEG from all patients are similar with respect to presence of burst-suppression patterns. Apparently there are characteristics of BS as detected by an experienced neurophysiologist that are not captured by the features selected here, no matter what combination used. In order to make the detection work equally well for all patients, other features or different classification techniques are needed.

In order to produce a complete burst-suppression detection- and quantization system, artifact rejection need to be included. It could also be useful to have a pre-classifier to first decide if the data is an example of burst-suppression or another type of periodic EEG activity, before the segmentation is made.

When more data is obtained, it will be possible to do a more statistically reliable training and evaluation of the neural network. The results of this will hopefully provide an easily used and reliable tool for automatic segmentation of burst-suppression EEG in newborns that has suffered from asphyxia. This tool may be a part of a monitoring system for neonatal intensive care units, as well as a tool for segmenting burst-suppression data used for off-line analysis.

References

[1] G. Holmes, J. Rowe, J. Hafford, R. Schmidt, M. Testa, and A. Zimmerman, "Prognostic value of the electroencephalogram in neonatal asphyxia," *Electroencephalography and Clinical Neurophysiology*, vol. 53, pp. 60-72, 1982.

[2] D. B. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. M. T. Robertson, "EEG and long-term outcome of term infants with neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology*, vol. 110, pp. 655-659, 1999.

[3] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology*, vol. 27, pp. 93-101, 2002.

[4] M. Thordstein, B. G. Wallin, A. Flisberg, R. Bågenholm, I. Kjellmer, N. Löfgren, and K. Lindecrantz, "Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates," *Clinical Neurophysiology*, vol. 115, pp. 2461-2466, 2004.

[5] M. D. Lamblin, M. André, M. J. Challamel, L. Curzi-Dascalova, A. M. D'Allest, E. De Giovanni, F. Moussalli-Salefranque, Y. Navelet, P. Plouin, M. F. Radvanyi-Bouvet, D. Samson-Dollfus, and M. F. Vecchierini-Blineau, "EEG in premature and full-term newborns. Maturation and glossary," *Neurophysiologie Clinique*, vol. 29, pp. 123-219, 1999.

[6] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology*, vol. 89, pp. 980-1002, 1998.

[7] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing*, vol. 15, pp. 357-367, 1999.

[8] C. Jeleazcov, S. Egner, F. Bremer, and H. Schwilden, "Automated EEG preprocessing during anaesthesia: New aspects using artificial neural networks," *Biomedizinische Technik*, vol. 49, pp. 125-131, 2004.

[9] J. A. Rice, *Mathematical Statistics and Data Analysis*, 2 ed. Belmont: Duxbury Press, 1995.

[10] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[11] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley Publishing Company, 1991.