Published in final edited form as: IEEE Trans Inf Technol Biomed. 2011 November ; 15(6): 839–847. doi:10.1109/TITB.2011.2159805.

EEG Signal Description with Spectral-Envelope-Based Speech Recognition Features for Detection of Neonatal Seizures

Andriy Temko [Member, EMBS],

Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork, Ireland. andreyt@eleceng.ucc.ie

Climent Nadeu,

Speech Processing Group, TALP Research Center, Department of Signal Theory and Communication, Univesitat Politècnica de Catalunya, Barcelona, Spain. climent.nadeu@upc.edu

William Marnane [Member, EMBS],

Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork, Ireland. I.marnane@ucc.ie

Geraldine Boylan, and

Department of Pediatrics and Child Health and the Neonatal Brain Research Group, University College Cork, Ireland. g.boylan@ucc.ie

Gordon Lightbody

Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork, Ireland. g.lightbody@ucc.ie

Abstract

In this work, features which are usually employed in automatic speech recognition (ASR) are used for the detection of seizures in newborn EEG. In particular, spectral envelope based features, composed of spectral powers and their spectral derivatives are compared to the established feature set which has been previously developed for EEG analysis. The results indicate that the ASR features which model the spectral derivatives, (either full-band or localized in frequency), yielded a performance improvement, in comparison to spectral-power based features. Indeed it is shown here that they perform reasonably well in comparison with the conventional EEG feature set. The contribution of the ASR features was analyzed here using the Support Vector Machines (SVM) recursive feature elimination technique. It is shown that the spectral derivative features consistently appear among the top-rank features. The study shows that the ASR features should be given a high priority when dealing with the description of the EEG signal.

Keywords

Neonatal Seizure Detection; EEG; Speech Recognition features; Spectral Envelope; Spectral Slope

I. INTRODUCTION

The task of automated detection of seizures in newborn EEG has undergone major advances in recent years. There have been two key directions of system development. The first one has focused on the creation of a set of heuristic rules and thresholds. Threshold based methods consist in analyzing the EEG using a small number of descriptors from which a decision is made using empirically derived thresholds. Recent examples of this methodology involve the work in [1-3]. The second approach relies on classifier based methods which

employ elements of pattern recognition to classify a set of features using a data-driven decision rule [4-6].

The main drawback of threshold-based approaches is that they result in a fixed single operating point, where there is little or no control over the trade-off between the good detection rate and the number of false detections per hour. In contrast, the employment of a classifier results in a continuous (probabilistic) output which allows the system to be easily adapted to the desired performance or to better meet the requirements of a particular application.

The choice of the best features is an important part of developing a good classification system especially when using the pattern recognition paradigm. There have been several studies on the importance and ranking of various features and feature sets for the task of neonatal seizure detection [4, 5]. As a result, a set of diverse features has been formed by including features which capture the frequency, energy, and structural content of the signal [6]. However, only few studies consider to a sufficient depth the feature sets which are conventionally used in other domains of signal processing. For instance, in the much more established area of audio/speech processing, features such as cepstral coefficients (CC) [7, 8] or frequency-filtered band energies (FF) [9, 10] are almost exclusively used, as they have been proven over a number of years to be capable of capturing all necessary information for that task. Unlike other spectral features used in EEG, the features used in ASR model the logarithmically scaled ratios of subband energies, which are either localized in frequency (FF) or full-band (CC). In the CC case, the features result from combinations of those energy ratios. Thus, the ASR features, either FF or CC, can be interpreted in terms of spectral derivatives, i.e. spectral slopes, in contrast to spectral powers. In fact, it has been shown in [28] that a phonetic distance based on spectral slopes correlates with perceptual speech data better than other speech characteristics such as filter-bank energies (FBE). Spectral slopes are less sensitive to details of the spectral shape and thus are more robust to inter- and intra-speaker variations. It is often mistakenly believed that ASR features based on the filter-bank energies, such as CC or FF, are matched to some underlying speech production model and hence they could hardly be employed for other types of signals. Actually, CC or FF features are just useful ways, based on linear transformations, of presenting to the classifier the information contained in the FBEs. The FBEs are a discrete representation of the smoothed spectral energy distribution, which is needed for both speech/ speaker recognition and EEG-based neonatal seizure detection. Actually, in both cases, a one dimensional signal is divided into segments (called frames in ASR or epochs in EEG signal processing), and a set of features is extracted to represent the characteristics of the signal in each segment. In both areas, characteristics extracted from the spectral envelope have been shown to be relevant for those tasks. The difference lies in the fact that, while in speech the frequency distribution of the sub-bands follows a perceptually-based quasilogarithmic scale, EEG on the other hand requires a uniform distribution. Whereas many studies [8-10] report superior performance of spectral slope features over other spectral features in speech and speaker recognition, further processing of the EEG spectral envelope has had limited investigation for biomedical signal applications, and indeed, contradictory conclusions could be drawn from the work that has been presented to date. Cepstral distance has been contrasted to spectral distance for EEG in adult rats, [11], where the authors showed that the results achieved using cepstral distance were superior to those achieved with spectral distance for the calculation of the brain injury index. In contrast, in [12] it has been shown that, despite the authors' expectations, cepstral processing did not improve the results for classification of rapid eye movement in an electromyography signal. Statistical parameters (such as mean, variance, kurtosis, etc) of cepstral coefficients extracted from an EEG epoch have been investigated in [13] for the detection of neonatal seizures. As a result of such statistical averaging within the EEG epoch, no features from cepstral analysis were

selected as being relevant for EEG classification. On the contrary, in [14], statistics calculated in time over the cepstral coefficients have been shown to be useful for the classification of the background EEG in neonates. It should be noted that this particular study only considered one set of spectral envelope parameters, the cepstral coefficients [14].

The work, presented here, extends our initial study on the application of speech recognition features for EEG signal description [15]. In particular, spectral envelope based features are compared to the well established feature set which was developed over years of research, for neonatal seizure detection [6, 16]. The results presented here indicate that spectral slope ASR features perform reasonably well on their own as a compact feature set arising from a single domain. Feature selection based on the SVM recursive feature elimination technique is utilized to determine how powerful these ASR features really are in comparison to the well established multi-domain EEG features. It is demonstrated that the spectral slope features consistently appear among the top-rank features, usually selected just after the corresponding spectral power features. The authors propose an explanation on why spectral slope features provide complementary information for the task of neonatal seizure detection and can also be seen as a more robust alternative to existing spectral power features.

The paper is organized as follows: Section II details the clinical dataset used in the experiments, introduces the investigated features, and describes the neonatal seizure detection system developed in the group, along with the feature selection routine. Section III presents and discusses obtained results. Conclusions are drawn in Section IV.

II. METHODS

A. Database

The dataset is composed of EEG recordings from 17 newborns obtained from the neonatal intensive care unit (NICU), Cork University Maternity Hospital, Cork, Ireland. The patients were full term babies ranging in gestational age from 39 to 42 weeks. All newborns had seizures secondary to hypoxic ischemic encephalopathy. A Carefusion NicOne video EEG machine was used to record multi-channel EEG at 256Hz using the 10-20 system of electrode placement modified for neonates. The data were annotated using 8 EEG channels in bipolar montage: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3 and C3-T3 (Fig. 1). All seizures were annotated independently by two experienced neonatal electroencephalographers with the assistance of simultaneous video recordings. All disagreements in annotations were resolved by consensus. The combined length of the recordings totals 267.9h and contains 705 seizures, which makes this dataset one of the largest in the neonatal area. The dataset contains a wide variety of seizure types including both electrographic-only and electro-clinical seizures of focal, multi-focal and generalized types. The EEG recordings were not edited to remove the large variety of artifacts and poorly conditioned signals commonly encountered in the NICU. Therefore this dataset is truly representative of the real-life NICU situation and its size allows the most robust estimate of the algorithm's performance. The dataset used is detailed in Table I.

B. Features

The EEG is down-sampled from 256Hz to 32Hz with an anti-aliasing filter set at 12.8Hz. It has been reported in [26] that the frequency range of neonatal seizures is between 0.5-13Hz with the dominant components in the 0.5-6Hz range. Neonatal seizures (Fig. 1) in EEG are defined as periods of increased periodicity for duration of over 10 seconds [27]. In our work, the EEG is split into 8s epochs with 50% overlap between epochs.

Unlike older children and adults, neonates do not always exhibit obvious clinical signs during seizures. Similarly, the difference between seizure and non-seizure EEG in neonates

can be relatively subtle, thus requiring a large set of features to quantify the change. A set of 55 features (Baseline, see Table II), which carry frequency, energetic and structural information (information theory) and which are previously exploited in [6, 16, 20], is extracted from the frequency and time domains. The majority of these features are taken from three studies on features for neonatal seizure detection. In [17], the authors performed a study on the value of chaos theory and information theory-based features. In [5], the importance and the performance of several popular features extracted from time-domain analysis have been compared. In [13], the authors used filter techniques for feature selection from a large set of primarily time domain-based features.

Apart from the Baseline feature set, 6 other sets of features are investigated here. Fig. 2 shows the block-diagram for calculation of these 6 spectral envelope feature sets:

Linear filter-bank energies (linFBE)—The spectrum is linearly smoothed using 15 triangular windows, with an overlap of 50%. Linear non-normalized powers are computed. Eleven of these features are actually already present in the Baseline feature set as shown in Table II. Here however, this set of 15 is considered as a stand-alone feature set for comparative purposes.

Relative filter-bank energies (relFBE)—These are linFBE energies normalized by the total power. Similarly, 11 of these features are already present in the Baseline feature. Again, this set of 15 is considered as a stand-alone feature set for comparative purposes.

Log filter-bank energies (logFBE)—Log scaling is applied to the 15 linFBE energies. The logFBE are the basic features considered in speech and speaker recognition. The nonlinear compression of the large amplitude range of spectral measurements has an additional advantage of converting any gain factor into an additive component which can be easily removed.

Cepstral coefficients (CC)—Cepstral coefficients [7, 8] are probably the most used spectral representation in speech. They are calculated by applying the discrete cosine transform (DCT), or the inverse Fourier transform to the log-magnitude Fourier spectrum. In speech recognition, the DCT is applied to the above-mentioned sequence of logFBEs. In fact, the DCT is almost equivalent to the Karhunen-Loève transform (also known as principal component analysis, PCA) for the sequence of logFBEs due to the fact that this sequence can be well approximated by a 1st order Markov chain and the value of the real pole of the model is close to 1 [9, 10]. It is also shown that these transforms (DCT and PCA) are asymptotically equivalent for an arbitrary wide sense stationary process [29]. Additionally, the DCT also sorts the transformed coefficients in order of increasing variance. The resultant vector is then truncated by retaining the highest energy components which can be seen as an implicit extra smoothing of the spectral envelope. All the cepstral coefficients except the zeroth one can be seen as combinations of spectral slopes that provide information about the rate of change in the different spectrum bands. Additionally, due to the periodic nature of the DCT basis functions, all the subbands are involved in their computation, so the CCs are not localized in a particular frequency. In this work, 15 cepstral coefficients were retained after application of the DCT to 20 log filter-bank energies. In the usual way [8], the zeroth cepstral coefficient was removed, and the epoch log energy was added to the set.

Frequency-filtered band energies (FF)—Spectral derivatives in the continuous frequency domain can be replaced by differences in values at discrete frequencies (when filter-bank analysis is performed, as in our case). This procedure can be achieved by

application of the second-order filter $H(z) = z - z^{-1}$ [9, 10] to the frequency sequence of logFBE. This derivative-type filter implies the subtraction of the logFBE of the two bands adjacent to the frequency band of interest. Before filtering, the logFBE sequence is extended along the frequency axis by the addition of an extra zero at each side. In this case, the first and last parameters actually represent the energies of the second and the penultimate subbands, and as such, the epoch log energy is not used with these features. Unlike CC, FF parameters are localized in frequency as they are obtained with explicit frequency filtering. They still lie in the frequency domain and preserve a frequency meaning. Here, 15 FF parameters are extracted from 15 logFBE.

Relative spectral difference (RSD)—It has been shown in [18], that the derivative of the spectrum can be calculated both before and after log-scaling:

$$\frac{d}{d\omega}S(\omega) = \frac{d}{d\omega}\log(E(\omega)) = \frac{1}{E(\omega)} \cdot \frac{d}{d\omega}E(\omega)$$

where $E(\omega)$ is the spectral envelope of the current EEG epoch, and $S(\omega)$ is its log-scaled version. The FF calculation can then be formulated as:

$$FF(k) = S(k+1) - S(k-1) = 2$$

where k is the frequency subband index. Similarly, RSD can be calculated as:

$$RSD(k) = \frac{E(k+1) - E(k-1)}{E(k+1) + E(k) + E(k-1)} \quad 3$$

where smoothing is introduced in the denominator of (3) as suggested in [18]. Thus, the RSD parameters can be seen as an alternative to the FF parameters, which avoids the computation of the logarithm.

These 6 feature sets attempt to retain coarse spectral energy distribution. The first three are power-based, while the last three are slope-based; either full-band (CC) or localized in frequency (FF, RSD).

Utilizing the above mentioned 6 features sets it is then possible to compare non-normalized powers (linFBE) with completely energy-independent powers (relFBE) and also with the logarithmically scaled powers (logFBE) which conceptually lie between the first two feature sets. The processing step required to calculate spectral slopes (CC, FF and RSD) from spectral powers (logFBE), involves a decorrelation transform (DCT or frequency filtering) which can be seen as a data independent alternative to the data-driven principal component analysis.

The average log spectra of both seizure and non-seizure classes can be seen in Fig. 3. Here cluster centers obtained using k-means clustering of both the seizure and non-seizure log-spectral envelopes are shown. It can be seen that seizure is characterized with slightly higher energy in every subband. The artifacts, which represent a small portion of the non-seizure data, are indicated with crosses. Due to the data imbalance which is common in neonatal seizure detection, the amount of artifact, although it may be negligible within the non-seizure class, can still be comparable to the amount of the seizure data. This fact decreases the discriminative capability of the energy-based features. On the other hand, the seizure spectral envelope is also characterized by the overall envelope slope, which is slightly larger for the seizure class than for the non-seizure class. This property can be described using the full-band slope features such as CC. Additionally, there is some activity in the first 4

subbands, which is visible in the seizure class and not present in the non-seizure class, and which can be described using the local slope features such as FF or RSD.

C. Neonatal Seizure Detection System

The neonatal seizure detection paradigm employed in this work is shown in Fig. 4. The training data for the classifier were first normalized anisotropically by subtracting the mean and dividing by standard deviation to assure commensurability of the various features. This normalizing template was then applied to the testing data. The normalized features extracted from each epoch were then used to train a single SVM classifier (with Gaussian kernel). The same classifier is used as in [6, 20] to assure the full comparability of the results.

Neonatal seizures are generally marked from the onset to the offset irrespective of which channels are involved. Per channel annotations are not produced as they do not add any additional information in neonates: a baby is equally treated if the seizure is on one or all channels. However, for the seizure class representation, per channel annotations are required to indicate on which channels the seizure event occurs and thus which channels should be used in training. Additional per-channel seizure annotations were produced for 2 minutes per patient of the entire dataset, which sum up to M^*2 minutes per patient for seizures involved in *M* channels. For example, if a training dataset consists of 16 patients for which 2 minutes of seizure are transcribed on the per-channel basis and on average 4 channels are involved in every seizure, then for an epoch length of 8 seconds with an overlap of 4 seconds, the seizure class of training data will approximately consist of 16patients*(120s/4s)*4channels = 1920 epochs. It may be more or less depending on the number of channels involved in every per-channel-annotated seizure for every patient. In our performance assessment, which is explained later in Section II.E, the number of seizure epochs ranged from a minimum of 1507 to a maximum of 1745, while testing patients 15 and 4, respectively. For the non-seizure class, 10,000 epochs of the non-seizure data were randomly selected from all channels of all patients in training dataset. Model selection on the training data was performed to choose suitable model parameters such as kernel hyper-parameters. Neonatal seizures can be localized to a single EEG channel (Fig. 1); for this reason the system was designed to process and classify each EEG channel independently. The outputs of the SVM were converted to pseudo probabilistic values using Platt's method [19] and smoothed with a moving average filter. The averaged value was then compared to a threshold from the interval [01]. After comparison, binary decisions were taken per channel: 1 for seizure and 0 for non-seizure. The binary decisions were then fused as follows: if there was a seizure in at least one channel, the whole epoch was marked as a seizure, otherwise it is denoted as a non-seizure. The 'collar' technique was applied last - every seizure decision was extended from either side to account for the delay introduced by the moving average filter and to compensate for possible difficulties in detecting pre-seizure and post-seizure sections. This system with the baseline feature set has thoroughly been compared to several existing alternatives and significantly outperformed them [20]. Detailed information on the seizure detection system and comparison results can be found in [6, 20].

D. Feature Selection Routine

Here, the widely used Recursive Feature Elimination (RFE) is employed to provide feature ranking to the SVM classifier. It was initially proposed in [21] for selecting genes that are relevant for a cancer classification problem. In this routine, an SVM classifier with Gaussian kernel is trained and the deviation in the cost function, DJ_i , is computed for every single feature that is removed, while preserving the same set of support vectors:

$$DJ_i = \left| \frac{1}{2} \alpha^T H \alpha - \frac{1}{2} \alpha^T H_i \alpha \right| \quad 4$$

Here *H* is the matrix with elements, $H(p, q) = y_p y_q K(x_p, x_q)$, $K(\cdot)$ is the radial basis function, x_p , x_q are support vectors and y_p , y_q are their respective labels. The set of support vectors with corresponding Lagrange multipliers (*a*) is assumed to remain unchanged for the H_i matrix which is re-computed without feature *i*. To relax this assumption, only a single feature is eliminated per iteration and the model is then retrained to balance the space of features and resultant space of support vectors. The feature corresponding to the smallest *DJ* is then removed. The RFE effectively provides nested subsets of features by sequentially eliminating each least useful feature. In our work, it is used to rank features as explained in the Experiments and Discussion section.

E. Performance Assessment and Metrics

Patient-specific data (labels) are usually not available to the detector in the clinical setting. Thus, it is important to retain this characteristic in the experimental setup. Furthermore, due to the limited availability of neonatal EEG data, it is important to maximize the use of data. For these reasons, the leave-one-out (LOO) cross-validation method was used here to assess the performance of the system for patient-independent seizure detection [20]. In this manner, all but one patients' data were used for training and the remaining patient's data were used for testing. This procedure was repeated until each patient had been a test subject and the mean result across all the patients was reported. The LOO method is known to be an almost unbiased estimation of the true generalization error [22].

The metric used in this work is the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve plots sensitivity and specificity values, which are defined as the accuracy of each class (seizure and non-seizure) separately.

III. EXPERIMENTS AND DISCUSSION

A. Performance of the Spectral Envelope Feature Sets

A comparison of the per-patient and average ROC areas for the original feature set (Baseline), 3 spectral power feature sets and 3 spectral slope feature sets is presented in Fig. 5.

The left plot of Fig. 5 shows that the most straightforward feature set, composed of linear spectral powers (linFBE) results in an ROC area of 89.3%. The relative spectral power feature set (relFBE), which is energy-independent and thus can be seen as the most robust of the 3 power feature sets, achieves the worst performance of 86.8%. This fact shows how robustness can be traded for information loss – by introducing energy-independence in the feature set, its discriminating capacity decreased. In particular, it is reflected in the lower performance shown on patients 2 and 10. The EEG of the patient 2 is vastly affected by different types of high energy artifacts (similar to shown in Fig.3). This is the reason why for patient 2 the performance with linear energy (linFBE) and the performance without energy (relFBE) are similarly low. On the contrary, the EEG of the patient 10 is relatively clean from artifacts and its performance with energy-independent features (relFBE) is lower than the performance with energetic feature sets (lineFBE, logFBE). At the same time, all seizures in this patient are quite subtle and as a result the performance with only powerbased features is not high. Among the 3 power-based feature sets, the best performance is obtained using logFBE which can be thought of as lying conceptually somewhere between linFBE and relFBE – the energy is not completely avoided but non-linearly scaled instead.

The performance of the system on slope feature sets is shown in the right plot of Fig. 5. Taking into account that these features were explicitly developed for audio/speech recognition and not for EEG signal description, the performance is surprisingly good. It can be seen from the right plot of Fig. 5 that all 3 spectral slope feature sets show similar behaviour for the task. The patients which resulted in lower performance for the baseline feature set also show a similar trend for the spectral slope feature sets. The only exception is patient 7 whose performance with the ASR features is decreased, whereas with the baseline feature set it is recovered by features that additionally carry other information apart from the spectral envelope.

Comparing the left and right plots it can be seen that further processing of the spectral envelope indeed resulted in both higher overall performance and significantly less interpatient performance deviation.

B. Feature Selection Results

In order to gain further insight into the role of the spectral envelope features, the baseline feature set was augmented with the 3 spectral envelope feature sets: logFBE, FF, and CC. The other spectral power feature sets (linFBE and relFBE) were not considered as they were already present in the Baseline. Likewise, the RSD spectral slope feature set was not included as it was deemed to carry similar information to the FF set. Thus, the RFE routine was applied to the 4 feature sets pooled together; that is 100 features. The mean ROC area for various numbers of preserved features is shown in Fig. 6.

Obviously, there is considerable information redundancy within this concatenated feature set. For instance, in ASR, the spectral envelope features are never used together.

Nevertheless, the results shown in Fig. 6 confirm that the SVM is insensitive to the presence of irrelevant features; this has also been shown in [23, 24]. Indeed, the SVM results with all 100 features together were not significantly worse than the results with a lower number of selected features. Additionally, it is also confirmed that omitting good features may be more detrimental for the SVM than including bad ones; this has also been shown in [24]. Here, with only around 10 features the ROC performance can be seen to have already reached 95%.

C. Feature Ranking

It is interesting to examine the order in which features are eliminated and try to find a smaller subset of features which would lead to a similar performance. In the obtained ranking, it is also interesting to observe the place of the spectral slope features when considered within the wider context. It is easier to show the importance of e.g. cepstral features if they are the only spectral envelope features considered, as was presented in [11]. The presence of the large number of other features assures that a feature is not emphasized due to the lack of alternatives that carry similar information but rather due to its higher robustness and discriminative power.

It is worth noting that Fig. 6 shows the performance of the system for feature sets of various lengths. However, because the RFE routine is applied to every training set in the LOO performance assessment, the ranking order of features (or the order in which features were eliminated) is different for every training dataset. That is, Fig. 6 shows the performance, averaged over all the testing patients, of the detector on a subset of say *N* features. However, these features may be different for every testing patient. It can be seen from Fig. 6 that the relative peak of performance lies between 20 and 50 features. To identify features that appear to be useful in all LOO cross-validation loops, the top N = 10, 20 30, 40, and 50

features, after the RFE routine was applied, were examined, and only the features that consistently appeared among the top N features for all training datasets were retained.

This simple procedure allows generalization of the importance of a particular feature. Indeed, a feature may be important for one patient and be useless for another, and as such this feature is unlikely to serve as a good descriptor for the task. The results of the feature ranking are summarised in Table III. As mentioned in the Methods section, the obtained feature subsets are nested. That is, for instance, the feature subset of 11 features, which was obtained by examining the top 20 features, also included the 3 features selected by examining the top 10 features. Thus, the last column of the Table III shows only the new features to be added.

From Table III it can be seen that in examining the top 10 features, there are 3 features which appear to be consistently important for all patients in the dataset: The log of total power (CC1); Shannon entropy; and the normalized power in the subband 1-3Hz. Among the top 20, there are 11 such features, which, in addition to those chosen from the top 10, include: Spectral entropy; spectral edge frequency at 95%; normalised power in subband 3-5Hz; the localized spectral slopes (FF2, FF3) which are the difference of the log powers in the triangular windows centred at 3Hz and 1Hz, and 4Hz and 2Hz, respectively; zero crossing rate of the first and second EEG time derivatives; and a full band spectral slope (CC2) which is a measure of the balance between the upper and lower halves of the spectrum. The feature set, which is selected from the top 50, already includes many higher order coefficients from FF and CC, which are concerned with increasingly finer features in the spectrum.

D. Feature Analysis

It can also be seen from Table III that spectral slope features represent a significant part of the final selected feature set accounting for 1 in 3 features, 4 in 11, 7 in 17, 10 in 22, and more than a half in the top 50: 19 out of the 34 common features. The representatives of the spectral envelope based features almost completely supplant frequency domain features of the Baseline feature set, such as the original sub-band powers, peak frequencies, or wavelet energy.

It is noted that both CC and FF feature sets globally share the same information (which is presented in different ways: FF is locally distributed, CC is not), and that their performance as distinct feature sets is also comparable (as seen in Fig. 5). However, it is interesting, that when choosing the features on an individual basis, the local slopes (FF) were preferred first. Looking at the ranked sequence of the selected features from the top 30, it can also be observed that power is typically followed by slope in the same band, that is: log total power (CC1) and full-band slope (CC2); normalized power 1-3Hz and slope 1-3Hz (FF2), normalized power 3-5Hz and slope 3-5Hz (FF4), etc. The only difference is the subband 2-4Hz, where the slope (FF3) precedes the power (normalized power 2-4Hz).

Looking at Fig. 3 it is not unexpected that low frequency components (powers and slope of the first 4 bands) prevail among the top-ranked features in Table III. Indeed, roughly a half of all features selected from the top 10, top 20, and top 30 are either powers or local slopes in the first 4 sub-bands. Regarding the full-band slopes, the first selected CC parameter after both the log total power (CC1) and the slope of the whole band (CC2), is CC10, which mainly measures the slope in the low frequency band (up to 3Hz). In fact, the contribution from the other subbands to CC10 may be neglected, since the spectral power has a decreasing tilt along frequency with a decay of around 25dB (as shown in Fig. 3).

From Table III it is also possible to observe that the normalized or scaled feature values are preferred over the absolute values – the relative subband powers are selected rather than linear subband powers, the selected CC (except CC1) and FF features are independent of the absolute signal energy, and additionally, the logarithmically scaled total power (CC1) is preferred over a linear total power.

It is reasonable to assume that improved performance can be obtained using a mixture of ASR features drawn from the ASR feature sets. Based on the observation that the first subband powers and slopes are usually included, a feature set containing the first five relFBE, the first five logFBE, the first 5 subband local slopes (FF2-FF6), the log total power (CC1), and the full-band slope (CC2) was formed and resulted in an ROC area of 94.0% (Sens=Spec=0.87, at the equal error rate point) which is higher than the ROC of any one ASR feature set alone.

It is interesting that information theory domain features occupy important positions in every subset. It is believed here that these features are mostly responsible for subtle but important differences in the performance between the spectral envelope features and the baseline shown in Fig. 5.

The performance of the system using the selected feature set was also assessed using the same LOO procedure with the detector trained on the fixed feature sets composed of 3, 11, 17, 22, and 34 features selected from top 10, 20, 30, 40, and 50, respectively, as explained in the previous section. In this manner, the feature set relevance was validated across all patients. Fig. 7 shows the 3D scatter plot of seizure and non-seizure epochs using the best 3 features. Although there is a substantial overlap between the classes, with these 3 best features the performance is already 92.8% (Sens=Spec=0.85). Adding the next 8 features is crucial – with 11 features, an ROC area of 95.4% (Sens=Spec=0.89) is achieved. A performance improvement of nearly 1% can be obtained by adding an extra 23 features (Top 50), eventually reaching similar performance to that of the baseline feature set of 55 features using in this case only 34 features (Sens=Spec=0.90).

The work indicates that the ASR features give a robust description of the EEG spectrum. It is suggested that the spectral slope features which have not been deeply exploited in EEG signal processing so far should be given a high priority. When the most informative part of the spectrum is known and targeted, then the localized slopes (FF and RSD) may be more beneficial than full-band slopes. Additionally, RSD provide an important benefit of avoiding the log calculation, which is essential if the system is to be implemented on an ambulatory device with built-in processor. An example is the REACT system on the Blackfin processor developed by this research group for the detection of seizures in adult EEG [25].

IV. CONCLUSIONS

This paper has demonstrated the importance of a proper investigation of other areas of signal processing for the task of EEG-based neonatal seizure detection. Several feature sets used in speech recognition/processing were exploited for the task. It has been shown that the imported spectral envelope based features provided reasonably good performance on their own, with spectral slope based features outperforming spectral power based features. Feature selection and ranking revealed that speech recognition features consistently ranked among the top selected features. Additionally, they almost completely masked the original EEG spectral features.

This work indicates that the spectral envelope based features can serve as a better and more robust alternative to the original spectral features used in the EEG signal description. Accompanied with several features from the information theory and time domains, ASR

features form a good representation of the signal for neonatal seizure detection and should be given a high priority when describing EEG signal in other applications.

Acknowledgments

This work was supported in part by the Welcome Trust (085249/Z/08/Z) and SFI (10/IN.1/B3036, 07/SRC/I1169).

Biography

Andriy Temko (S'05-M'10) received the Engineering degree in Informatics in 2002 from Dniepropetrovsk National University, Dniepropetrovsk, Ukraine and the PhD degree in Telecommunication in 2008 from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. His main research interests include kernel methods, signal processing, and multimodal interfaces. During 2006-2007 he was a task leader in detection and classification of acoustic events within the EU-funded international evaluation campaigns on detection of events, activities, and their relationships (CLEAR 2006/CLEAR 2007). Since late 2008 he has been with Neonatal Brain Research Group, University College Cork, Ireland, working on algorithms for EEG and ECG based seizure detection in newborns babies and adults. He has been involved in several EU and national governments funded projects on audio/speech and biomedical signal processing.



Climent Nadeu received the Telecommunication Engineering degree in 1977 and the Doctoral degree in 1982, both from the Universitat Politècnica de Catalunya (UPC), Barcelona. Since 1977 he has been with the UPC where he is Professor on signal processing (from 1991). During his sabbatical leaves, he has been a visiting researcher at AT&T Bell Laboratories, Murray Hill (NJ), at the International Computer Science Institute, Berkeley (CA), and at Griffith University, Brisbane (Australia). He has more than 180 publications in books, scientific journals and conference proceedings, mainly in the area of speech technologies. He is member of the Editorial Board of the Speech Communication journal, and Associated Editor of the Journal on Audio, Speech and Music Processing.



William Marnane received the B.E. degree in electrical engineering from the National University of Ireland, Cork, in 1984, and the Ph.D. degree from the University of Oxford, Oxford, U.K., in 1989. He was a lecturer at the School of Electronic Engineering Science, University of Wales, Bangor from 1989 to 1993. In 1993 he was appointed as a Lecturer in Digital Signal Processing in the Department of Electrical & Electronic Engineering at University College Cork and as a Senior Lecturer in 1999. His research interests include Biomedical Signal Processing and digital design for DSP, coding and cryptography.



Geraldine B. Boylan received the M.Sc. degree in physiology and the Ph.D. degree in clinical medicine from University College London, London, U.K. She worked as a Clinical Scientist in Neonatal Medicine in Kings College Hospital London from 1996–2001. She is currently a Senior Lecturer in Paediatrics in the Department of Paediatrics & Child health, University College Cork, Cork, Ireland. Her research interests concentrate on accurately diagnosing seizures in newborn babies by monitoring EEG and studies of blood flow regulation during neonatal seizures. Much of her more recent work is of an interdisciplinary nature and aims to create a synergy between medicine and engineering by using the skills and techniques of engineering signal processing research to address important medical problems such as seizure detection in the neonate.



Gordon Lightbody graduated with the MEng degree (distinction) (1989), and then PhD (1993) both in Electrical and Electronic Engineering from Queen's University Belfast. After completing a one year Post-Doctoral position funded by Du Pont, he was appointed by Queen's University as a Lecturer in Modern Control Systems. In 1997 he was appointed as a Lecturer in Control Engineering at University College Cork, and subsequently to a Senior Lecturer in 2008. His current research interests include artificial intelligence techniques for intelligent control and signal-processing, focusing on biomedical and energy/power applications. He is a member of the IET, and is currently an associate editor with the Elsevier journal, "Control Engineering Practice".

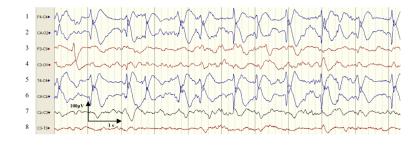


REFERENCES

- Mitra J, Glover J, Ktonas P, Kumar A, Mukherjee A, Karayiannis N, Frost J, Hrachovy R, Mizrahi E. A Multistage System for the Automated Detection of Epileptic Seizures in Neonatal Electroencephalography. J Clin Neurophysiol. 2009; 26:1–9. [PubMed: 19151615]
- [2]. Deburchgraeve W, Cherian P, Vos MD, Swarte R, Blok J, Visser G, Govaert P, Huffel SV. Automated neonatal seizure detection mimicking a human observer reading EEG. Clin Neurophysiol. 2008; 119:2447–54. [PubMed: 18824405]

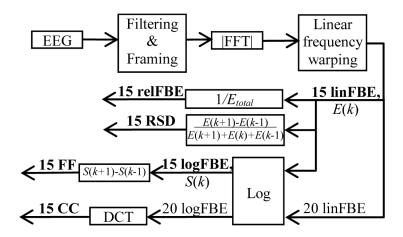
- [3]. Navakatikyan M, Colditz P, Burke C, Inderd T, Richmond J, Williams C. Seizure detection algorithm for neonates based on wave-sequence analysis. Clin Neurophysiol. 2006; 117:1190– 1203. [PubMed: 16621690]
- [4]. Aarabi A, Grebe R, Wallois F. A multistage knowledge-based system for EEG seizure detection in newborn infants. Clin Neurophysiol. 2007; 118:2781–97. [PubMed: 17905654]
- [5]. Greene B, Faul S, Marnane W, Lightbody G, Korotchikova I, Boylan G. A comparison of quantative EEG features for neonatal seizure detection. Clin Neurophysiol. 2008; 119:1248–61.
 [PubMed: 18381249]
- [6]. Temko A, Thomas E, Boylan G, Marnane W, Lightbody G. EEG-Based Neonatal Seizure Detection with Support Vector Machines. Clin Neurophysiol. 2011; 122:464–73. [PubMed: 20713314]
- [7]. Kemerait R, Childers D. Signal Detection and Extraction by Cepstrum Techniques. IEEE T Inform Theory. 1972; 18:745–59.
- [8]. Rabiner, L.; Juang, BH. Fundamentals of Speech Recognition. Prentice-Hall; Englewood Cliffs, NJ: 1993.
- [9]. Nadeu C, Hernando J, Gorricho M. On the decorrelation of filter-bank energies in speech recognition. Proc. Eurospeech. 1995
- [10]. Nadeu C, Macho D, Hernando J. Time and frequency filtering of filter-bank energies for robust HMM speech recognition. Speech Commun. 2001; 34:93–114.
- [11]. Geocadin RG, Ghodadra R, Kimura T, Lei H, Sherman DL, Hanley DF, Thakor NV. A novel quantitative EEG injury measure of global cerebral ischemia. Clin Neurophysiol. 2000; 111:1779–87. [PubMed: 11018492]
- [12]. Shokrollahi M, Krishnan S, Jewell D, Murray B. Autoregressive and Cepstral Analysis of Electromyogram in Rapid Movement Sleep. Proc. IFMBE. 2009; 25:1580–1583.
- [13]. Aarabi A, Wallois F, Grebe R. Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis. Clin Neurophysiol. 2006; 117:328–240. [PubMed: 16376606]
- [14]. Löfhede J, Thordstein M, Löfgren N, Flisberg A, Rosa-Zurera M, Kjellmer I, Lindecrantz K. Automatic classification of background EEG activity in healthy and sick neonates. J Neural Eng. 2010; v.7:1–12.
- [15]. Temko A, Boylan G, Marnane W, Lightbody G. Speech Recognition Features for EEG Signal Description in Detection of Neonatal Seizures. Proc. IEEE EMBC. 2010:3281–3284.
- [16]. Thomas E, Temko A, Lightbody G, Marnane W, Boylan G. Gaussian mixture models for classification of neonatal seizures using EEG. Physiol Meas. 2010; v.31:1047–1064. [PubMed: 20585148]
- [17]. Faul, S.; Boylan, G.; Connolly, S.; Marnane, W.; Lightbody, G. Chaos theory analysis of the newborn EEG. Is it worth the wait?. Proc. IEEE Int. Symp. on Intelligent Signal Processing; 2005. p. 381-386.
- [18]. Nadeu, C. On the filter-bank-based parameterization front-end for robust HMM speech recognition. Proc. Workshop on Robust Methods for Speech Recognition in Adverse Conditions; 1999.
- [19]. Platt J. Probabilistic outputs for SVM and comparison to Regularized likelihood methods. Advances in Large Margin Classifiers. 1999
- [20]. Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. Performance assessment for EEGbased neonatal seizure detectors. Clin Neurophysiol. 2011; 122:474–482. [PubMed: 20716492]
- [21]. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002; 46:389–422.
- [22]. Vapnik, V. Estimation of Dependences Based on Empirical Data. Springer-Verlag; New York: 1982.
- [23]. Weston J, Mukherjee J, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. Proc. NIPS. 2000:668–674.
- [24]. Guyon I, Aliferis C, Cooper G, Elisseeff A, Pellet J, Spirtes P, Statnikov A. Design and Analysis of the Causation and Prediction Challenge. J Mach Learn Res. 2008; 3:1–33.

- [25]. Temko, A.; McEvoy, R.; Dwyer, D.; Faul, S.; Lightbody, G.; Marnane, W. REACT: Real-Time EEG Analysis for Seizure Detection; Proc. AMA-IEEE Medical Tech. Conf. on Individualized Healthcare; 2010.
- [26]. Kitayama M, Otsubo H, Parvez S, Lodha A, Ying E, Parvez B, Ishii R, Mizuno-Matsumoto Y, Zoroofi RA, Snead OC. Wavelet analysis for neonatal electroencephalographic seizures. Pediatr Neurol. 2003; 29:326–323. [PubMed: 14643396]
- [27]. Shellhaas R, Clancy R. Characterization of neonatal seizures by conventional EEG and singlechannel EEG. Clin Neurophysiol. 2007; 118:2156–2161. [PubMed: 17765607]
- [28]. Klatt H. Prediction of perceived phonetic distance from critical-band spectra: A first step. Proc. IEEE ICASSP. 1982:1278–1281.
- [29]. Unsera M. On the approximation of the discrete Karhunen-Loeve transform for stationary processes. Signal Processing. 1984; 7:232–249.



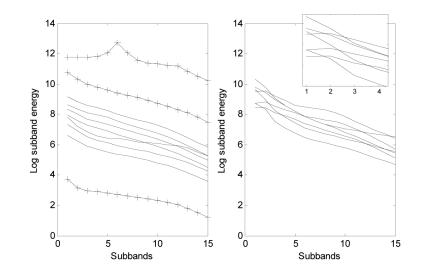


Example of a neonatal seizure - a focal seizure localized in the right central region, C4 (channels 1, 2, 5, 6).





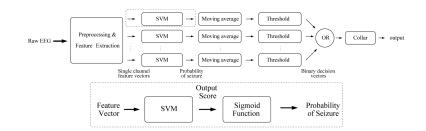
Block diagram for calculation of spectral envelope features. The 6 feature sets used are indicated in bold.

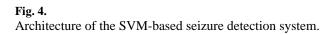




Cluster centers of the log spectra for non-seizure (left) and seizure (right) classes. The lines with crosses on the left plot indicate cluster centers with low data density, which correspond to artifacts.

Europe PMC Funders Author Manuscripts





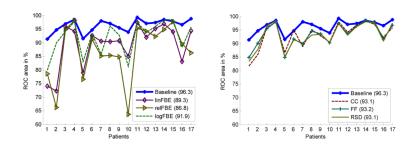
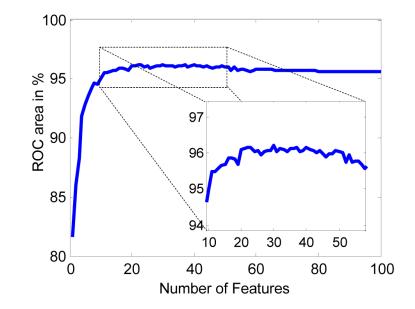


Fig. 5.

Performance of the SVM-based seizure detection system. Left plot shows the comparison of the per-patient and average ROC areas for 3 spectral power feature sets (linFBE, relFBE, logFBE). Right plot shows the comparison of the per-patient and average ROC areas for 3 spectral slope feature sets (CC, FF, RSD).





Recursive feature elimination applied to the set of 100 features which includes Baseline, logFBE, CC, and FF feature sets. The x-axis shows the number of preserved features.

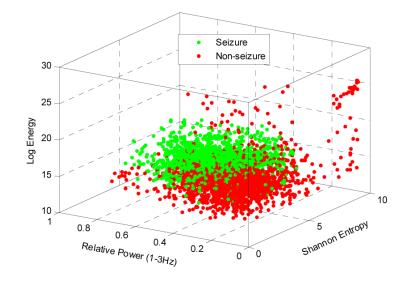


Fig. 7.

Seizure and non-seizure epochs in the 3D space composed of 3 best features – Log total energy (CC1), Shannon Entropy, and relative power (1-3Hz).

TABLE I

EEG DATASET

Patient	Record length (h)	Seizure events
1	18.23	17
2	24.74	3
3	24.24	149
4	26.10	60
5	24	49
6	5.69	41
7	24.04	6
8	24.53	17
9	24.04	156
10	10.06	25
11	6.19	15
12	12	29
13	12.13	25
14	5.48	11
15	12.16	59
16	7.63	31
17	6.64	12
Total	267.9	705

TABLE II

FEATURES EXTRACTED FOR EACH EPOCH

Analysis	Features	
	 Total power (0-12Hz), - Peak frequency of spectrum, Spectral edge frequency (SEF80%, SEF90%, SEF95%), Power in 2Hz width subbands (0-2Hz, 1-3Hz,10-12Hz), - Normalised power in same subbands, - Wavelet energy (Db4 wavelet coefficient corresponding to 1-2Hz) 	
Baseline (55)	- Curve length, - Number of maxima and minima, - Root mean square amplitude, - Hjorth parameters (activity, mobility and complexity), - Zero Crossing Rate (ZCR), - ZCR of the Δ and the $\Delta\Delta$, - Variance of Δ and $\Delta\Delta$, - Autoregressive modelling error (AR model order 1-9), - Skewness, - Kurtosis, - Nonlinear energy	
	Shannon entropy, - Spectral entropy, - Singular Value Decomposition entropy, - Fisher information	
linFBE	15 subband energies (0-2Hz, 1-3Hz,)	
relFBE	15 subband energies normalised by total energy	
logFBE	15 logarithmically scaled subband energies	
CC	15 cepstral coefficients	
FF	15 second order frequency filtered bank energies	
RSD	15 relative spectral difference	

TABLE III

FEATURE RANKING. UNDERLINED ARE FEATURES WHICH COME FROM THE ASR FEATURE SETS

Range	# Top Ranked (from ASR)	Features
Top 10	3 (1)	CC1, Shannon entropy, Normalised power in sub-band 1-3Hz
Top 20	11 (4)	Spectral Entropy, SEF95, Normalised power in sub-band 3-5Hz, <u>FF2</u> , <u>FF3</u> , ZCR of the Δ and the $\Delta\Delta$, <u>CC2</u>
Top 30	17 (7)	ZCR, FBE3, Normalised power in sub-band 2-4Hz, <u>FF7, FF4, FF6</u>
Top 40	22 (10)	FF8, FF9, Kurtosis, Normalised power in sub-band 8-10Hz, <u>CC10</u>
Top 50	34 (19)	CC5, AR modelling 1, <u>FF11</u> , FBE1, Skewness, <u>FF10</u> , <u>CC12</u> , <u>CC13</u> , Fisher information, <u>CC4</u> , <u>CC11</u> , <u>FF5</u>