Artefact detection in neonatal EEG*

N.J. Stevenson[1], J.M. O'Toole[1], I. Korotchikova[1], and G.B. Boylan[1]

Abstract— **Artefact detection is an important component of any automated EEG analysis. It is of particular importance in analyses such as sleep state detection and EEG grading where there is no null state. We propose a general artefact detection system (GADS) based on the analysis of the neonatal EEG. This system aims to detect both major and minor artefacts (a distinction based primarily on amplitude). As a result, a two-stage system was constructed based on 14 features extracted from EEG epochs at multiple time scales: [2, 4, 16, 32]s. These features were combined in a support vector machine (SVM) in order to determine the presence of absence of artefact. The performance of the GADS was estimated using a leave-one-out cross-validation applied to a database of hour long recordings from 51 neonates. The median AUC was 1.00 (IQR: 0.95-1.00) for the detection of major artefacts and 0.89 (IQR: 0.83-0.95) for the detection of minor artefacts.**

I. Introduction

The automated analysis of neonatal EEG is an emerging field of research. These automated methods of analysis are designed to provide important information on cortical function to the clinician, in the neonatal intensive care unit, with the intention of improving clinical management of this vulnerable population. The requirement for continuous analysis of a complex signal, around the clock, has resulted in the application of computer assisted methods. To date, there have been several systems that attempt to classify states within the EEG such as seizure, sleep state and various other abnormalities [2], [3], [5].

A state that is common to all these systems is a 'missing data' state which defines the absence of EEG of sufficient quality to classify clinically relevant states. The primary cause of the reduction in EEG quality is artefact. Artefacts are generated by electrical or biological processes and contaminate the EEG signal making the visual, and, therefore, computational interpretation of the underlying EEG difficult.

There has been surprisingly little research dedicated solely to the detection of artefacts in the neonatal EEG as early development of automated methods tend to assume that the EEG under analysis is artefact free. This precludes practical implementation of such algorithms until their robustness can be tested. Methods that have achieved a level of robustness have circumvented the problem of artefacts by assigning artefact to a null case such as the non-seizure state in the seizure detection task [5]. The problem with such an approach is that the detection of seizure should be predicated on an assessment of the quality of the underlying EEG, as EEG that is contaminated with artefact cannot be used to ascertain the presence or absence of a seizure. It would be preferable if this 'missing data' state were highlighted. There are also several automated EEG tasks such as sleep state detection or EEG grading that do not have a null state with which artefact can be bundled. In these cases, an additional 'missing data' state would complement the automated analyses.

While there have been an array of methods developed for the detection and removal of EEG artefacts [4], [6], there has been little effort made to develop stand-alone artefact detection systems for neonatal EEG that have been tested on large, clinically relevant datasets.

In this paper, we propose a system for general artefact detection in the neonatal EEG. The GADS is proposed within a machine learning paradigm and is based on the support vector machine (SVM). The novelty of the proposed system is the multi-stage implementation which detects minor and major artefact separately and the feature extraction processes which estimates features on epochs at multiple time scales in order to contextualise a short duration epoch with information from a longer duration epoch. The performance of the each stage of the GADS was estimated on a cohort of 51 neonates using a leave-one-out cross-validation.

II. Data Acquisition

The data were acquired from the neonatal intensive care unit of the Cork University Maternity Hospital (CUMH), Ireland, using a Nicolet One multiple channel video-EEG system. A total of 51 term neonates were monitored with the EEG due to a suspected neurological injury. A total of 9 electrodes were placed according to the international 10-20 system and an 8 channel bipolar montage: F4-C4, C4-02, C4-Cz, C4-T4, F4-C4, C3-01, Cz-C3, C3-T3, was used for analysis. The data were recorded with a sampling frequency of 256Hz and filtered with a band pass filter [0.5, 70]Hz with an additional 50Hz notch filter (the detection of 50Hz artefact was not a priority as notch filtering provides adequate removal). Artefacts within the data were annotated by an experienced electroencephalographer. The annotations were not channel specific and merely denoted the presence of artefact on at least a single channel of the EEG recording. A large array of artefacts were annotated including cardiac, respiratory, body, muscle, electrode contact, and high impedance (electrode disconnect) [1]. A summary of the database is shown in Table I and example artefacts from 6 neonates are shown in Fig. 1. A total of 55h of EEG recording from 38 neonates was used to generate the minor

*This work was supported by Science Foundation Ireland (12-IP-1369) and the Qatar National Research Foundation (NPRP:4-1303-2-517)

[1] Neonatal Brain Research Group, Irish Centre for Fetal and Neonatal Translational Research, University College Cork, Ireland n.stevenson at ucc.ie

978-1-4244-7929-0/14/$26.00 ©2014 Crown

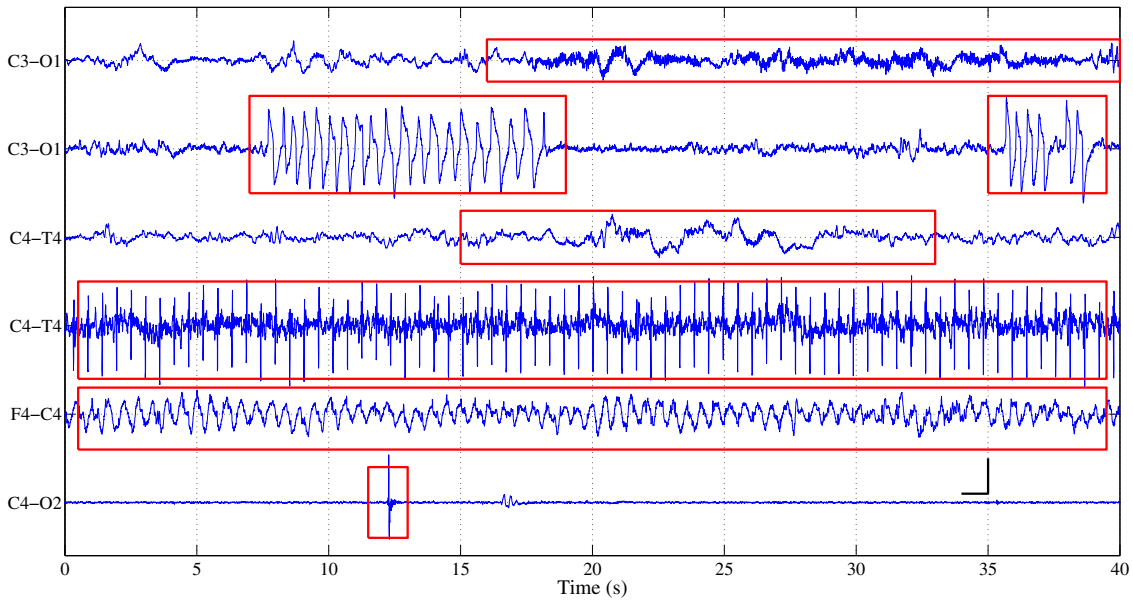

Fig. 1. The manifestations of artefact on a single channel of neonatal EEG from 6 different neonates. Artefacts are highlighted by the red box. From top to bottom: muscle, electrode, movement, heart rate, respiration and transient electrode. The voltage scale is equal for all examples, the black marker at 35 s is 100 μV.

TABLE I

SUMMARY OF THE DATABASE USED TO DEVELOP THE ARTEFACT DETECTION SYSTEM. VARIABLES ARE SUMMARISED ACROSS THE COHORT OF 38 AND 13 NEONATES FOR THE MINOR AND MAJOR ARTEFACT DATABASES, RESPECTIVELY.

	Minor	Major
Subject Number	38	13
Recording Duration (min)	70 (65-76)	139 (118-157)
Artefact No	36 (21-88)	13 (13-13)
Artefact Duration (s)	4.8 (2.5-7.5)	4384 (3664-4632)

artefact database and a total of 35h of EEG recording from 13 neonates was used to generate the major artefacts database.

All EEGs were recorded with written, informed, parental consent and under ethical approval of the CUMH and the University College Cork. All data were anonymized at the time of recording.

III. GENERALIZED ARTEFACT DETECTION SYSTEM

A two-stage patient-independent GADS was proposed. The first stage aimed to differentiate large amplitude artefacts such as those caused by electrode disconnect or poor impedance between the electrode and the skin. The second stage aimed to detect more subtle manifestations of artefact such as muscle, movement and periodic artefacts. The outputs of both stages were combined using a logical OR operation as a final post-processing step. An overview of the GADS is shown in Fig. 2.

The two stages of the GADS were proposed within a standard machine learning paradigm, that is, a sequence of features estimated on an epoch of neonatal is passed to a classifier based on a SVM. The output of the classifier was then used to determine the presence of absence of artefact by applying a simple threshold. Several pre-processing and post-processing stages were added to condition the incoming data and improve the decision output. Pre-processing steps included the implementation of a high pass filtering with a single pole infinite impulse response filter with a cutoff frequency 0.5Hz, a 50Hz notch filter, and segmentation into epochs of various duration around a central time point). Post-processing steps included taking the maximum SVM output across 8 EEG channels which was then filtered with a mean filter of 6s duration and a collaring operation that extends the detection 2s forwards and backwards in time.

There are many features used for the analysis of neonatal EEG [4], [7]. We selected a small subset of features that were assumed to respond to different types of artefact based on a preliminary visual analysis of the artefacts in the database. The 14 features used are the mean, median and variation of the amplitude, mean frequency, bandwidth, three frequency-band energies, a ratio of maximum energy to mean energy, a ratio of the maximum absolute value of the fractional derivative to the mean energy, the Bera-Jarque statistic estimated on the fractional derivative and an estimate of the Hurst exponent. These features are selected as they respond to both the presence of background EEG and artefact. Two features were specifically implemented for the detection of repetitive artefacts such as ECG and respiration: the peak frequency and spectral distortion.

The mean and variance of the amplitude of the signal were

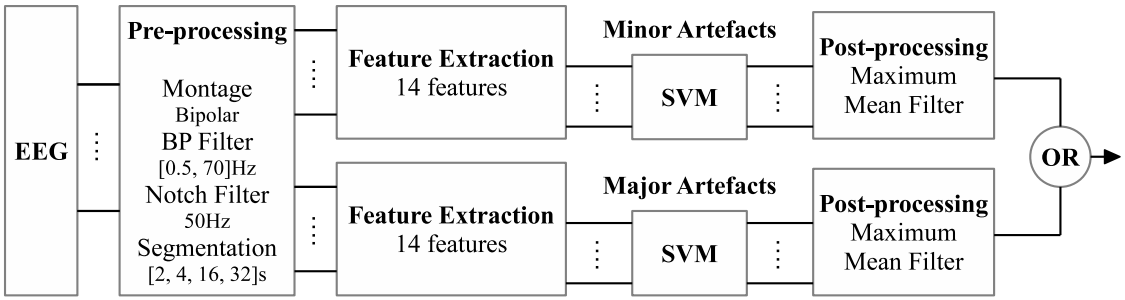

Fig. 2. The general artefact detection system.

estimated using the analytic associate of a signal,

$$F_1 = \frac{1}{N}\sum_{n=0}^{N-1} |\mathrm{eeg}(n) + j\mathcal{H}\{\mathrm{eeg}(n)\}| \quad (1)$$

$$F_2 = \mathrm{median}(|\mathrm{eeg}(n) + j\mathcal{H}\{\mathrm{eeg}(n)\})| \quad (2)$$

$$F_3 = \frac{1}{N}\sum_{n=0}^{N-1} (|\mathrm{eeg}(n) + j\mathcal{H}\{\mathrm{eeg}(n)\}| - F_1)^2 \quad (3)$$

where $\mathrm{eeg}(n)$ is the epoch of EEG under analysis, $\mathcal{H}$ is the discrete Hilbert transform of a signal, N is the discrete epoch length, $n = 0, ..., N-1$, f_s[2, 4, 16, 32]s, and $f_s = 256$Hz.

The mean frequency and bandwidth were defined as,

$$F_4 = \frac{\sum_{k=0}^{N/2} nP(k)}{\sum_{k=0}^{N/2} P(k)} \quad (4)$$

$$F_5 = \frac{\sum_{k=0}^{N/2} (n-F_4)^2 P(k)}{\sum_{k=0}^{N/2} P(k)} \quad (5)$$

where $P(k)$ is the periodogram of the signal epoch $\mathrm{eeg}(n)$, and k is discrete frequency, $k = 0, ..., N-1$.

Additional measures of band energy were also included as diagnostically relevant information in the neonatal EEG is predominantly located in the delta (0.5-4 Hz),theta (4-8 Hz), and alpha (8-13 Hz) bands. This results in three measures which estimate the signal energy in each frequency band,

$$F_6 = \sum_{k=0.5/K}^{13/K} P(k) \quad (6)$$

$$F_7 = \frac{\sum_{k=0}^{0.5/K} P(k)}{F_6} \quad (7)$$

$$F_8 = \frac{\sum_{k=13/K}^{70/K} P(k)}{F_6} \quad (8)$$

where $K = f_s/N$. The estimation of F_6 to F_8 were performed on the EEG without the pre-processing filter.

The fractional derivative of the EEG signal was estimated by deconvolving a filter $h(n)$. This filter is defined as,

$$h(n) = \begin{cases} 1, & n = 0 \\ \frac{(H+0.5+n)h(n-1)}{(n+1)}, & n = 1, ..., N-1 \end{cases} \quad (9)$$

where H is the Hurst exponent (estimated using the wfbmesti function in Matlab; the wavelet based estimate of H is used). The Hurst exponent was also used as a feature (F_9).

Two ratios between quantiles (maximum and median) of the absolute values of the EEG signal and its fractional derivative are also used.

$$F_{10} = \frac{\max(|\mathrm{eeg}(n)|)}{\mathrm{median}(|\mathrm{eeg}(n)|)} \quad (10)$$

$$F_{11} = \frac{\max(|\mathrm{eeg}(n) *^{-1} h(n)|)}{\mathrm{median}(|\mathrm{eeg}(n) *^{-1} h(n)|)} \quad (11)$$

where $*^{-1}$ is the deconvolution operation. The deconvolution operation can be performed as the magnitude of the frequency domain representation of the filter is greater than zero at all frequencies.

The Bera-Jarque test statistic was defined as,

$$F_{12} = \frac{N}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \quad (12)$$

where N is the signal length, S is the skewness and K is the kurtosis. This statistic is used as neonatal EEG can be assumed to be amplitude modulated fractional Brownian process [3]; a property of such a process is that it should have a Gaussian increment.

Finally, measures of peak frequency and spectral distortion were defined as,

$$F_{13} = \mathrm{argmax}_k P(k) \quad (13)$$

$$F_{14} = \frac{P(F_{13})}{\sum_{k=0}^{N/2} P(k)} \quad (14)$$

The application of these features to EEG analysis is not novel; in fact several are analagous to the features used in [4]. The novelty of this feature set, however, is that it applies the same features to EEG epochs segmented at different time scales: 2s, 4s, 16s, and 32s. The major artefact detection stage uses 32s epochs only and the minor artefact detection stages used 2s, 4s, and 16s epochs. The short duration epoch localises the detection and epochs of longer duration contextualises the information over longer time scales. This attempts to mimic the annotation of the neurophysiologist where the recording will be scanned both before and after suspect EEG activity in order to make a decision as to the presence or absence of artefact.

A. Training and Testing

The performance of the GADS was estimated using a leave-one-subject-out (LOSO) cross-validation. A maximum of 50 epochs per class from each neonate were used to form the training set. This resulted in 493 and 320 mins of single channel EEG data at each training step (based on the maximum epoch length used in each stage). The median duration of testing was 70 and 139 mins of 8-channel EEG for the minor and major artefact databases, respectively.

All features were bounded on $[0, B]$ where $B < \infty$, and were transformed using the natural logarithm and then converted to z-scores; the former provides little benefit in training but aids visualisation during development.

The performance of the GADS was assessed with the area under the receiver operator characteristic (AUC). The receiver operator characteristic is generated by evaluating the sensitivity (percentage of seconds correctly identified by the GADS as contaminated by artefact based on the visual interpretation of the electroencephalographer) and specificity (percentage of seconds correctly identified by the GADS as not contaminated by artefact based on the visual interpretation of the electroencephalographer) over a range of thresholds bounded by the extrema of the SVM output.

IV. Results and Discussion

The results of the LOSO cross-validation are shown in Table II and Table III. Table II outlines the performance summarized across all possible thresholds with the AUC and Table III outlines several event and time based measures (see [5]) evaluated at a single threshold. This threshold corresponds to a false alarm rate of 6/h and 0.1/h for minor and major artefact detection, respectively (differences in thresholds reflect differences in the duration of minor and major artefacts). The statistics are summarised across 38 (minor dataset) and 13 (major dataset) neonates.

The detection of major artefacts is trivial as there are significant differences in amplitude and frequency between background EEG and artefact. This is reflected in the results.

The detection of minor artefact is a more difficult proposition. The short duration of the majority of minor artefacts is the primary reason for this difficulty as there is less data available to estimate the features. In fact, the performance of the GADS is comparable to other methods of automated EEG analysis (such as neonatal seizure detection [5]) when assessed without post-processing; the mean AUC without post-processing was 0.86 and 0.85 for the seizure detection algorithm and minor artefact detection system, respectively. The improvements achieved in seizure detection from post-processing the SVM output are not as apparent in the detection of minor artefact as the mean event duration is significantly less: 249 s for neonatal seizure as opposed to 5 s for minor artefact. There are also several other factors which impede the performance of patient-independent detection of minor artefacts in the NICU: large intra- and inter-patient variability in background EEG patterns and voltages (isoelectric to hypsarrhythmia) and uncertainty in visual annotation of artefacts (expert agreement has been shown to have an AUC of 0.95 for visual interpretation for a limited set of artefacts in adult sleep [4]).

TABLE II

The performance of GADS across all detection thresholds.

	Median (IQR)	Mean
AUC (Major 32s only)	1.00 (0.95-1.00)	0.98
AUC (Minor)	0.89 (0.83-0.95)	0.85
AUC (Minor 2s only)	0.86 (0.78-0.93)	0.82
AUC (Minor 4s only)	0.88 (0.78-0.92)	0.83
AUC (Minor 16s only)	0.87 (0.76-0.93)	0.82

TABLE III

The performance of GADS at a single detection threshold.

	Major	Minor
Sensitivity (%)	89.1 (64.4-100.0)	62.5 (25.1-79.1)
Specificity (%)	98.7 (98.3-100.0)	96.3 (93.1-97.7)
Artefact Detection Rate (%)	100.0 (100.0-100.0)	62.9 (39.3-78.4)

There is clearly room for improvement when detecting minor artefacts in the neonatal EEG. A significant factor neglected in the current system is the multi-channel nature of EEG. Certain types of artefact are more likely to have a presence on multiple channels simultaneously. The classifier and/or post-processing stage can be modified to incorporate multi-channel data and several features that measure the synchrony and symmetry between channels provide future avenues of investigation.

V. Conclusion

A system for detecting major and minor artefacts on the neonatal EEG proposed. The performance of the GADS was compared to the visual interpretation of the human expert via a LOSO cross-validation and had a median AUC across subjects of 1.00 for major and 0.89 for minor artefact detection stages. The performance of the GADS was increased when features from epochs at multiple time scales were classified.

References

[1] G.B. Boylan, D.M. Murray, & J.M. Rennie, "The normal EEG and aEEG", in eds. J.M. Rennie, C. Hagmann, & N. Roberston, *Neonatal Cerebral Investigations*, Cambridge, UK: Cambridge University Press, 2008.

[2] A. Piryatinska, G. Terdik, W. A. Woyczynski, K. A. Loparo, M. S. Scher, and A. Zlotnik, Automated detection of neonate EEG sleep stages, Comput. Meth. Prog. Bio., vol. 95, pp. 31-46, July 2009.

[3] N.J. Stevenson, I. Korotchikova, A. Temko, G. Lightbody, W.P. Marnane, G.B. Boylan, An automated system for grading EEG abnormalities in term neonates with hypoxic ischaemic encephalopathy, Ann. Biomed. Eng., vol. 41, pp. 775-785, April 2013.

[4] P.J. Durka, H. Klekowicz, K. J. Blinowska, W. Szelenberger, and Sz. Niemcewicz, A simple system for detection of EEG artifacts in polysomnographic recordings, IEEE T. Bio-med. Eng., vol. 50, pp. 526-528, April 2003.

[5] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, EEG-based neonatal seizure detection with support vector machines, Clin. Neurophysiol., vol. 122, pp. 464-476, March 2011.

[6] S-Y Shao, K-Q Shen, C. J. Ong, E. P. V. Wilder-Smith, X-P Li, Automatic EEG artifact removal: A weighted support vector machine approach with error correction, IEEE T. Bio-Med. Eng., vol. 56, pp. 336-344, February 2009.

[7] B.R. Greene, S. Faul, W.P. Marnane, G. Lightbody, I. Korotchikova, and G.B. Boylan, A comparison of quantitative EEG features for neonatal seizure detection, Clin. Neurophysiol., vol. 119, pp. 1248-1261, June 2008.