Medical Engineering and Physics 45 (2017) 42–50

Contents lists available at ScienceDirect

Medical Engineering and Physics

journal homepage: www.elsevier.com/locate/medengphy

Detecting bursts in the EEG of very and extremely premature infants using a multi-feature approach

John M. O'Toole [a,*], Geraldine B. Boylan [a], Rhodri O. Lloyd [a], Robert M. Goulding [a], Sampsa Vanhatalo [b], Nathan J. Stevenson [a,1]

[a] Neonatal Brain Research Group, Irish Centre for Fetal and Neonatal Translational Research (INFANT), University College Cork, Ireland
[b] Department of Clinical Neurophysiology, Children's Hospital, HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

ARTICLE INFO

Article history:
Received 28 July 2016
Revised 27 March 2017
Accepted 2 April 2017

Keywords:
Burst detection
Electroencephalography
Preterm infant
Feature extraction
Spectral analysis
Support vector machine
Inter-burst interval

ABSTRACT

Aim: To develop a method that segments preterm EEG into bursts and inter-bursts by extracting and combining multiple EEG features. *Methods:* Two EEG experts annotated bursts in individual EEG channels for 36 preterm infants with gestational age $<$ 30 weeks. The feature set included spectral, amplitude, and frequency-weighted energy features. Using a consensus annotation, feature selection removed redundant features and a support vector machine combined features. Area under the receiver operator characteristic (AUC) and Cohen's kappa (κ) evaluated performance within a cross-validation procedure. *Results:* The proposed channel-independent method improves AUC by 4–5% over existing methods ($p < 0.001$, $n = 36$), with median (95% confidence interval) AUC of 0.989 (0.973–0.997) and sensitivity–specificity of 95.8–94.4%. Agreement rates between the detector and experts' annotations, $\kappa = 0.72$ (0.36–0.83) and $\kappa = 0.65$ (0.32–0.81), are comparable to inter-rater agreement, $\kappa = 0.60$ (0.21–0.74). *Conclusions:* Automating the visual identification of bursts in preterm EEG is achievable with a high level of accuracy. Multiple features, combined using a data-driven approach, improves on existing single-feature methods.

1. Introduction

Preterm birth is the single largest risk factor for perinatal mortality and morbidity, accounting for over 1 million deaths every year [1]. The immature brain of the preterm infant is especially vulnerable and often the source of long-term health problems. The electroencephalogram (EEG) can help identify at-risk infants by providing continuous cot-side monitoring of brain activity in the neonatal intensive care unit (NICU). The EEG, however, requires interpretation by specialist staff which often makes it impractical to provide continuous reporting for all infants. Automated EEG analysis could overcome this limitation and provide the clinician with relevant information, in real time, to guide treatment during critical care.

Early preterm EEG exhibits an intermittent or discontinuous pattern (*tracé discontinu*) consisting of low-voltage activity, known as inter-bursts, followed by short-duration higher-voltage activity, known as bursts or spontaneous activity transients [2]. This pattern differs to the burst-suppression pattern found in the EEG of adults and full-term infants, a pattern associated with severe brain injury or coma [3]. In contrast, the discontinuous pattern is indicative of normal, healthy neurological development for the preterm infant. An important first stage for any automated analysis of preterm EEG is to distinguish between bursts and inter-bursts. Simple features of this bursting pattern, such as maximum inter-burst duration, relate to neurological development and are associated with neurological delay [4–7]. Segmentation of the EEG into bursts and inter-bursts is an essential first-stage for more advanced automated analysis; for example to predict neurodevelopmental outcome [8], detect changes in sleep states [9], or assess changes in maturation [7].

Existing methods for detecting bursts in preterm EEG rely on either amplitude or frequency characteristics, or combinations of both [2,6,8,10–19]. Many of these methods, however, were not designed as stand-alone detection methods and have not been assessed with the gold standard, the EEG expert's visual interpretation of the EEG [2,8,10,11,13,16]. For those methods with performance validation metrics, the more promising methods employ frequency-weighted energy measures, which

* Corresponding author.
E-mail addresses: JOToole@ucc.ie, j.otoole@ieee.org (J.M. O'Toole), G.Boylan@ucc.ie (G.B. Boylan), RLloyd@ucc.ie (R.O. Lloyd), R.Goulding@ucc.ie (R.M. Goulding), Sampsa.Vanhatalo@Helsinki.fi (S. Vanhatalo), Nathan.Stevenson@Helsinki.fi (N.J. Stevenson).
1 Present address: BaBa Centre, University of Helsinki, Finland

http://dx.doi.org/10.1016/j.medengphy.2017.04.003
1350-4533/

multiply amplitude and frequency to estimate energy [6,17–19]. Yet the relative importance of amplitude and frequency features is unknown, and their optimal combination is as yet unexplored.

Here, we propose to assess multiple amplitude and frequency features separately and then combine these features in a classifier. This approach has been applied to detecting burst-suppression patterns in full-term EEG [20,21]. Based on training from a large database of preterm EEG, machine learning algorithms can infer the best combination rules. We apply a feature selection procedure, that maximises relevancy and minimises redundancy, thus retaining only necessary features. Unlike existing methods, which either operate on 1 specific channel [17] or all channels simultaneously [6,18], channels are processed independently as bursts can be focal or multi-focal and not always generalised across all channels. For example, in asynchronous activity bursts will not occur simultaneously across hemispheres [22]. For performance testing, feature sets and all parameters are estimated using strata of cross-validations to avoid overlap between training and testing data.

2. Methods

2.1. Acquiring and annotating the EEG

EEG data were collected from the NICU of the Cork University Maternity Hospital, Ireland, during the period 2009–2011. Data collection was approved by the Cork Research Ethics Committee of Cork Teaching Hospitals, Ireland. Informed and written parental consent was obtained before EEG recording.

EEG was recorded with the NicoletOne EEG system (CareFusion Co., San Diego, USA) using 11 electrodes according to the international 10–20 system of electrode configuration over the frontal, central, temporal, and occipital regions, a reference electrode at Fz, and a ground electrode behind the left ear. EEGs were recorded within 72 h of birth with a sampling frequency of 256 Hz. Infants with reported severe brain injuries, determined by cranial ultrasound scans within the first week of life, were not included.

Ten-minute segments with minimal artefact were selected from 36 EEG records (one segment per infant). These 10 min segments were, on average, 14 h post-birth (range: 3–41 h). Gestational age ranged from 23.4 weeks to 29.7 weeks with a mean of 27.4 weeks.

Two clinical physiologists (RO Lloyd and RM Goulding) annotated all EEG segments for bursts and inter-bursts. Bursts were defined as any preterm EEG activity not explicitly categorised as inter-bursts. Therefore the annotations included long-duration bursts ($>$ 20 s) which some classification systems would label as continuous activity [4]. We chose not to distinguish between bursts and continuous activity because the difference between continuous and discontinuous activity is not clearly defined for infants with gestational age less than 32 weeks [4]. Example annotations are in Fig. 1.

EEG was analysed using the bipolar montage F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, and C3-T3. EEG channels were annotated separately to develop a channel independent detector. As bursts do not always occur synchronously across all channels, a single channel was extracted for review to avoid annotation bias caused by the simultaneous display of multiple channels. One channel per infant was annotated and channel selection was alternated over all EEG records to avoid a channel bias. For example, F4-C4 was used for the first EEG, C4-O2 was used for the second, and so on. For all 36 EEGs, each channel was selected a median of 4.5 (range: 3–6) times.

Annotations differed between the two reviewers, as the example in Fig. 1 highlights. A consensus annotation, including only the burst or inter-burst periods where both reviewers agreed, was used for training and testing the classifier.

2.2. Feature set

Fig. 2 highlights differences between bursts and inter-bursts. For example spectral power, across all frequencies, is greater for bursts comparative to inter-bursts [Fig. 2(a)]. Not surprising, considering amplitude plays a key role in many detection methods [2,6,8,12,17–19].

But also of interest are spectral characteristics independent of total power. Differences in relative spectral power is evident in the normalised spectra in Figs. 2(b) and the burst-to-inter-burst ratio (the difference in spectral power in dBs between the median burst and inter-burst spectra) in Fig. 2(c). Fig. 2(b) shows that the inter-bursts have an almost linear log–log frequency response compared with the more nonlinear response of the bursts. The following feature set aims to capture these differences in amplitude, relative spectral power, and spectral shape. These features are calculated within four frequency bands: band 1 (0.5–3 Hz), band 2 (3–8 Hz), band 3 (8–15 Hz), and band 4 (15–30 Hz) [2,23].

2.2.1. Amplitude features. Discrete EEG signal $x(n)$ was bandpass filtered using a 5th-order Butterworth filter into the ith frequency band ($i = 1, 2, 3, 4$) to produce $x_i(n)$. These filters implement the forward–backwards procedure to produce a zero-phase filter. We calculated signal envelope $a_i(n)$ of $x_i(n)$ as

$$a_i(n) = |z_i(n)|^2 = |x_i(n) + j\mathcal{H}[x_i(n)]|^2 \quad (1)$$

where $z_i(n)$ is the analytic associate of $x_i(n)$; $\mathcal{H}$ represents the Hilbert transform and j represents the imaginary unit of the complex-valued $z_i(n)$.

2.2.2. Spectral features. Multiple features are used to quantify spectral characteristics. Relative spectral power for the ith band is estimated as

$$P_i = \frac{\sum_{k\in i} |X(k)|^2}{P_{\text{total}}} \quad (2)$$

where $X(k)$ is the discrete Fourier transform (DFT) of length-N $x(n)$, P_{total} is the total spectral power over the 0.5–30 Hz range, and notation $\sum_{k\,\in\,i}$ represents summation over the ith frequency band.

To quantify spectral shape, we fit the line

$$\hat{Y}(k) = c_1 + c_2 k \quad (3)$$

to the log–log spectrum $Y(k)$ and then use slope c_2 and measure-of-fit r^2, defined as

$$r_i^2 = 1 - \frac{\sum_{k\in i} \left[Y(k) - \hat{Y}(k)\right]^2}{\sum_{k\in i} \left[Y(k) - \frac{1}{N}\sum_{k\in i} Y(k)\right]^2}, \quad (4)$$

as features. This process has some similarity to a multifractal approach [24] but differs in the EEG frequency-band selection and summary measures.

Mean frequency is calculated using the periodic-mean frequency estimate,

$$M_i = \frac{f_s}{4\pi}\left\{\arg\left[\sum_{k=0}^{N/2-1} \left|X_i(k)\right|^2 e^{j2\pi k/N}\right] \mod 2\pi\right\} \quad (5)$$

with mod 2π representing the modulus function in 2π, f_s the sampling frequency, and $X_i(k)$ is the DFT of $x_i(n)$. Instantaneous frequency is calculated using the central-finite difference estimate,

$$f_i(n) = \frac{f_s}{4\pi}\left\{[\phi_i(n+1) - \phi_i(n-1)] \mod 2\pi\right\} \quad (6)$$