Journal of Neuroscience Methods 218 (2013) 110–120

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

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

Computational Neuroscience

Multimodal detection of head-movement artefacts in EEG

Simon O'Regan*, William Marnane

Department of Electrical and Electronic Engineering, University College Cork, Ireland

HIGHLIGHTS

- EEG artefacts are detected with SVM classifiers trained on EEG and gyroscope data.
- Combining EEG and gyroscope classifiers can improve artefact detection accuracy.
- An analysis of data fusion methods at feature and classifier levels is carried out.
- Feature and score level (sum rule) fusion are the best performing fusion methods.

ARTICLE INFO

Article history:
Received 26 October 2012
Received in revised form 17 April 2013
Accepted 20 April 2013

Keywords:
Electroencephalography
Artefact detection
Movement artefacts
Gyroscopes
Support vector machines
Classifier combination
Data fusion
Brain–computer interface

ABSTRACT

Artefacts arising from head movements have been a considerable obstacle in the deployment of automatic event detection systems in ambulatory EEG. Recently, gyroscopes have been identified as a useful modality for providing complementary information to the head movement artefact detection task. In this work, a comprehensive data fusion analysis is conducted to investigate how EEG and gyroscope signals can be most effectively combined to provide a more accurate detection of head-movement artefacts in the EEG. To this end, several methods of combining these physiological and physical signals at the feature, decision and score fusion levels are examined. Results show that combination at the feature, score and decision levels is successful in improving classifier performance when compared to individual EEG or gyroscope classifiers, thus confirming that EEG and gyroscope signals carry complementary information regarding the detection of head-movement artefacts in the EEG. Feature fusion and the score fusion using the sum-rule provided the greatest improvement in artefact detection. By extending multimodal head-movement artefact detection to the score and decision fusion domains, it is possible to implement multimodal artefact detection in environments where gyroscope signals are intermittently available.

1. Introduction

Ambulatory electroencelephalography (AEEG) is a valuable tool in a number of emerging medical applications. Automated seizure detection, widely accepted as useful in aiding a clinician in diagnosing patients with suspected epilepsy, can be extended to the domestic environment with the use of AEEG (Waterhouse, 2003; Casson and Rodriguez-Villegas, 2011). In brain–computer interface (BCI) applications such as cognitive state estimation, AEEG holds the potential to monitor tiredness levels in vehicle drivers and robotic surgery operators (Lan et al., 2007; Müller et al., 2008). Similarly, BCI-controlled text editors for disabled persons (Allison et al., 2007) can be extended to the mobile domain using AEEG (Lotte et al., 2009). However, contamination of the EEG signal by electrical artefacts arising from head-movements has been widely acknowledged as being problematic in ambulatory EEG (Waterhouse, 2003; O'Regan et al., 2012; Winkler et al., 2011; Gwin et al., 2010). By corrupting the EEG, head-movement artefacts can obscure the signal and interfere with its interpretation by a clinician/researcher. In automated neurological event detection systems, such as automated epileptic seizure detection or automated Alzheimer's disease recognition, head-movement artefacts may cause the classifier to falsely misinterpret a section of artefactual EEG as a neurological event (Kelleher et al., 2010; Lehmann et al., 2007).

Head movements can introduce a wide range of non-cerebral electrical activity into the EEG, typically taking the form of some combination of electrode pop, muscle (EMG), electrode movement and ocular artefacts. These component artefact signals display a wide range of characteristics. Electrode pop, which occurs when an electrode temporarily breaks contact with the surface of the scalp, is usually accompanied by fast, high amplitude spikes (Barlow, 1986). Muscle artefacts are predominantly high frequency signals, and can range from low to high amplitude (Willis et al., 1993; Brunner et al.,

* Corresponding author. Tel.:+353 21 4903156.
E-mail addresses: simonor@rennes.ucc.ie, simoregan@gmail.com (S. O'Regan).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.04.017

1996; Goncharova et al., 2003). Electrode movement traditionally results in slow-wave baseline drifts, but can sometimes manifest as apparent oscillation in the EEG. Ocular artefacts, introduced due to relative movement between the eye and the electrode, typically cause high amplitude deflections in the EEG (Lins et al., 1993; Berg and Scherg, 1991; Croft and Barry, 2000; Gasser et al., 1992; Picton et al., 2000). While these component artefact signals exhibit diverse temporal, frequency and structural characteristics, they are significantly different from those of background EEG. In previous work, it was shown that these component artefact signals (EMG, electrode pop, movement and ocular artefacts) could be grouped together in distinguishing them from normal EEG activity (O'Regan et al., 2010a,b, 2012).

Recent advances in the miniaturisation of gyroscopes have resulted in their inclusion in a number of commercially available EEG headsets (Emotiv EPOC, 2012). By measuring angular rotation of the headset, these gyroscopes represent a promising method of accurately detecting the occurrence of head-movements. As head movements often result in EEG artefacts (and certain head-movements may be more likely than others to introduce artefacts to the EEG), in previous work, the use of gyroscopes for artefact detection was introduced illustrating that features extracted from gyroscope signals are useful in determining if head-movement artefacts have contaminated the EEG (O'Regan et al., 2010b, 2012). The question then arose as to whether gyroscope features provide complementary information to features extracted from the EEG. By combining features extracted from the EEG and the gyroscope signals in a single classifier (feature fusion), it was found that artefact classification performance improved, suggesting that the gyroscope features contain information regarding the production of artefacts that is not provided by the EEG features alone.

Fusion at the feature level (sometimes referred to as early integration) is only one method of combining information from multiple signals for use in making a classification decision. Fusion at the classifier level (sometimes referred to as late integration) is a well-researched method of combining information from different modalities, with a range of options available to combine a set of classifiers into a final decision rule. Classifier output scores (score fusion) or binary output decisions (decision fusion) are two such classes of classifier fusion that allow individual classifiers to be combined to produce an overall classifier output. Classifier fusion offers a number of implementation advantages over feature fusion, such as more robustness in the absence of one of the input signals and easy incorporation of additional signals to the classification task.

There are many examples in the literature where classification using a combination of EEG and additional signals outperforms the base classifiers in the set (Peng et al., 2007; Kapoor et al., 2008; Qian et al., 2009; Polikar et al., 2010). However, the results reported in the literature do not clearly indicate whether early or late integration will offer better performance for the multimodal head-movement artefact classification task. In the seizure detection domain, Greene et al. (2007) investigated the combination of EEG and ECG signals in improving the performance of neonatal seizure detection and found feature fusion to outperform classifier fusion. Malarvili and Mesbah (2008) found the opposite, with fusion of EEG and ECG signals at the classifier level offering better detection of seizure in neonatal EEG. Bermudez et al. (2007) performed a similar analysis for detecting temporal lobe epilepsy in adults but found that classifier fusion offered better classification performance than feature fusion. In this paper, a number of methods combining EEG and gyroscope artefact detection classifiers are investigated. By taking advantage of both EEG and gyroscope signals, it is shown that the classification performance improvement introduced with feature fusion (early integration) in O'Regan et al. (2012) can be extended to classifier fusion (late integration). In doing so, the complimentary nature of EEG and gyroscope modalities regarding head-movement artefact is confirmed. Amongst the classifier fusion methods investigated, the score fusion sum-rule exhibited the best performance and offers a number of advantages to feature fusion; most notably, robustness in the face of intermittent gyroscope signals and easy integration of additional physiological signals should they become available. To the best of the authors' knowledge, this is the first time that an analysis of multimodal fusion of EEG and gyroscope signals has been conducted.

2. Materials and methods

2.1. Data collection

Subjects. 30 min of head movement data was collected from 7 healthy male adults (23–50 years, mean age 30). None of the participants had a history of neurological or psychiatric disorders and none were on chronic medication. Informed consent was obtained from all participants.

Artefact generation protocol. An artefact generation protocol was drawn up which instructed the participants to perform repetitions of each of the following movements: shake head, clench jaw, nod, roll head, raise and lower eyebrows. Between repetitions participants were asked to remain still in order to generate reference EEG. Particular focus was placed on movement artefacts that have been observed to occur more regularly in an ambulatory EEG system. The artefact generation protocol is described in detail in Table 1. Demonstrations of example movements were performed before recording took place, and the participants were instructed to perform similar movements at the designated times. Participants were instructed to perform the movements as naturally as possible and to vary the pace and direction of head movements where appropriate in order to avoid excessively repetitive, periodic artefacts that may be unlikely to occur in a natural ambulatory environment.

Experimental set-up. The 14-channel Emotiv EPOC EEG headset was used, sampled at 128 Hz (Emotiv EPOC, 2012). A referential montage was utilised, with reference electrodes taken from behind the ears. The Emotiv EPOC employs gold-plated contact-grade hardened copper electrodes with saline-moistened felt pads to record the EEG. This device has recently been used as an

Table 1
Artefact generation protocol.

Head-movements	Duration	Description
Shake head	30 s	Shake head from side to side, varying pace and direction
Remain still	20 s	Remain seated, avoiding head movements and eye blinks and movements
Clench jaw	30 s	Prolonged as well intermittent clenches (mimicking chewing)
Remain still	20 s	Remain seated, avoiding head movements and eye blinks and movements
Nod head	30 s	Nod head up and down, changing pace as doing so
Remain still	20 s	Remain seated, avoiding head movements and eye blinks and movements
Roll head	30 s	Roll head in both directions, changing pace as doing so
Remain still	20 s	Remain seated, avoiding head movements and eye blinks and movements
Raise and lower eyebrows	30 s	Changing pace and amplitude throughout
Remain still	20 s	Remain seated, avoiding head movements and eye blinks and movements

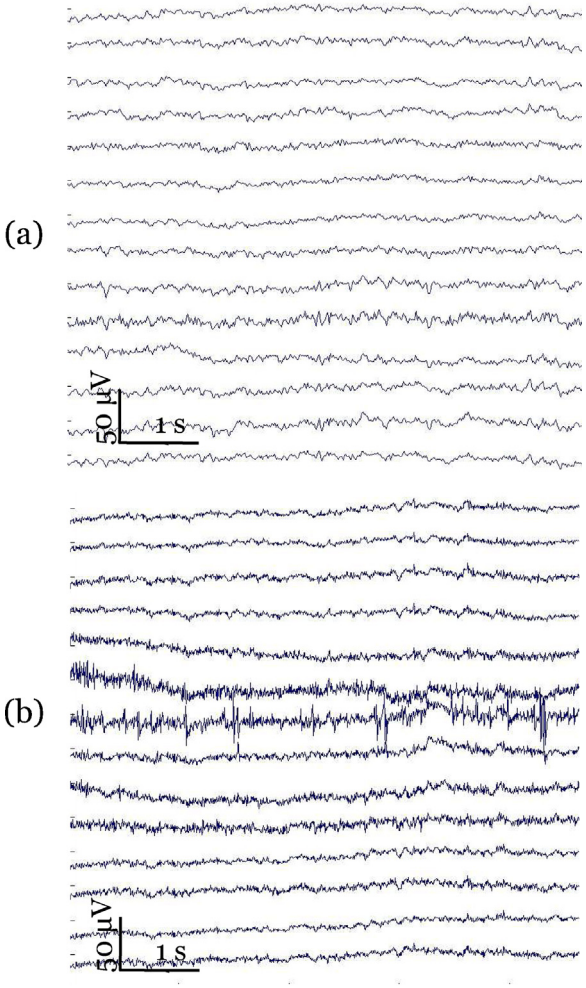

Fig. 1. Comparison of EEG recorded from the same participant with the Emotiv EPOC and EEG recorded using a clinical EEG recording device. (a) 5 s of background EEG recorded with the Emotiv EPOC, sampled at 128 Hz. (b) 5 s of background EEG from a clinical EEG recording using the Viasys NicoletOne, sampled at 250 Hz.

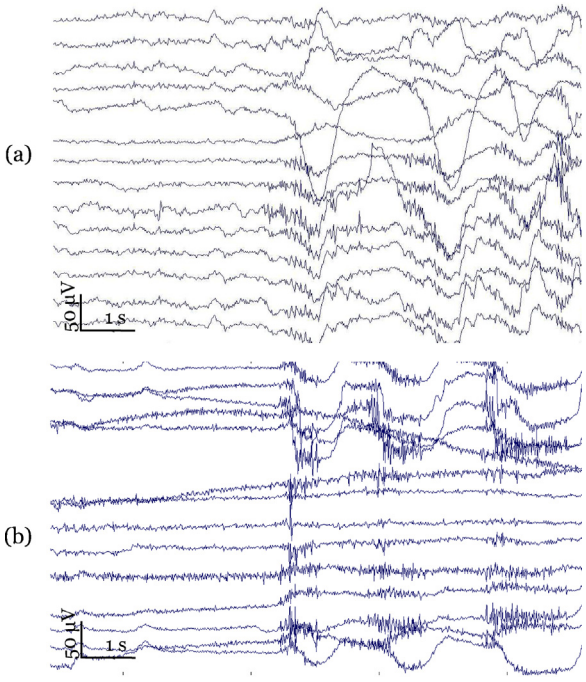

Fig. 2. Comparison of head-movement artefact EEG recorded from the same participant with the Emotiv EPOC and EEG recorded using a clinical EEG recording device. (a) 5 s of artefact EEG generated by nodding the head and recorded with the Emotiv EPOC, sampled at 128 Hz. (b) 5 s of artefact EEG generated by nodding the head using a clinical EEG recording using the Viasys NicoletOne, sampled at 250 Hz.

inexpensive substitute for medical EEG for cognitive neuroscience and BCI applications in the ambulatory domain (Anderson et al., 2011; Bobrov et al., 2011; Debener et al., 2012; Khushaba et al., 2013). Duvinage et al. (2012) report that although the signal quality in the Emotiv EPOC is lower than an equivalent medical-grade EEG system, it is suitable for use in BCI experiments. A comparison of EEG from the 128 Hz Emotiv EPOC and EEG from a clinical EEG machine, the Viasys NicoletOne sampled at 250 Hz are displayed in Figs. 1 (background EEG) and 2 (head-movement artefact). Gyroscope signals from the 2-axis gyroscope located at the rear of the headset were also sampled at 128 Hz. For both signals, the amplitude resolution was 0.51 μV per bit.

Annotation. The EEG artefact data was annotated as artefact where visually noticeable deflection in the EEG was observed at the times that participants performed movements. Annotation was independently confirmed by a specialist clinical physiologist in neurophysiology.

2.2. Detection system

The EEG and gyroscope classifiers implemented in this paper are based on previous work (O'Regan et al., 2012). These classifiers, as well as the feature, decision and score fusion classifiers share a number of common elements. The following paragraphs explain each of their stages in more detail.

Preprocessing. A fourth order Chebyshev notch filter was applied to the raw EEG and gyroscope data in order to remove the 50 Hz mains component from the signals. An additional fourth order, Butterworth highpass filter with cut-off at 1 Hz was applied to remove the DC-component.

EEG feature generation. In the EEG classifier (Fig. 3), EEG from each channel is separately pre-processed and segmented into overlapping windows. In order to maximise the amount of data available, overlap between windows of 90% was chosen. A set of features (outlined in Table 2) was extracted from each EEG window. The feature set was compiled to account for temporal, frequency and structural differences between background and artefact EEG. Features that have been demonstrated to perform well for the classification of the artefact sources created when head-movement occurs were chosen. Features describing EMG, ocular, electrode pop and movement artefacts (van de Velde et al., 1999; Delorme et al.,

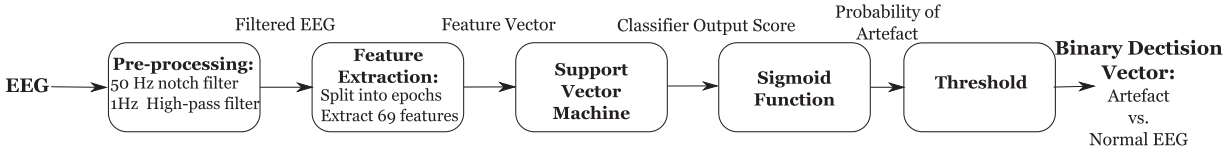

Fig. 3. System architecture for EEG artefact detection. This process is repeated for each EEG channel in parallel.

Table 2
List of features.

EEG features	**Gyroscope features**
Time domain features	*Time domain features*
•Line length/curve length (L)	•Zero crossings
•RMS Amplitude	•Nonlinear energy (NLE)
•Slope	•Line length/curve length (L)
•Activity (1st Hjorth parameter)	•Maximum value of autocorrelation
•Mobility (2nd Hjorth parameter)	•No. of inactive samples
•Complexity (3rd Hjorth parameter)	•RMS amplitude
•Kurtosis	•Activity (1st Hjorth parameter)
•Skewness	•Mobility (2nd Hjorth parameter)
•Nonlinear energy (NLE)	•Complexity (3rd Hjorth parameter)
•Zero Crossings (Zc)	•Kurtosis
•No. of inactive samples	•Skewness
•Autoregressive modeling error (AR models 1–9)	•Zero crossings of first derivative
•Variance of first derivative	•Zero crossings of second derivative
•Variance of second derivative	•Variance of first derivative
•Zero crossings of first derivative	•Variance of second derivative
•Zero crossings of second derivative	
Frequency domain features	*Frequency domain features*
•Peak frequency	•Total power
•Spectral edge frequency (80, 90 and 95)	
•Intensity weighted mean frequency (IWMF)	
•Intensity weighted bandwidth (IWBW)	
•Total power	
•Power bands: 0–2 Hz, 1–3 Hz, 2–4 Hz, 3–5 Hz, 4–6 Hz, 5–7 Hz, 6–8 Hz, 7–9 Hz, 8–10 Hz, 9–11 Hz, 10–12 Hz, 3–15 Hz, 15–30 Hz, 59–61 Hz, 51–64 Hz, 20–30 Hz, 25–64 Hz	
•Normalised power bands: 0–2 Hz, 1–3 Hz, 2–4 Hz, 3–5 Hz, 4–6 Hz, 5–7 Hz, 6–8 Hz, 7–9 Hz, 8–10 Hz, 9–11 Hz, 10–12 Hz, 3–15 Hz, 15–30 Hz, 59–61 Hz, 51–64 Hz, 20–30 Hz, 25–64 Hz	
Entropy-based features	
•SVD entropy	
•Shannon entropy	
•Fisher entropy	
•Spectral entropy	

2001; Gasser et al., 2005; Barlow, 1986) were supplemented with features from EEG seizure detection (Thomas, 2011) that showed visual separation between artefact and normal EEG. In O'Regan et al. (2010a) mutual information evaluation function and linear discriminant analysis were implemented to evaluate the effectiveness of using these features for the head-movement artefact classification task. In previous work, the feature fusion classifier performed best for window lengths of 1 s. Therefore, in this experiment, a window length of 1 s was investigated.

Gyroscope feature generation. The Emotiv EPOC headset provides two gyroscope signals, *x*-direction movement and *y*-direction movement; where *x*-direction is lateral rotation around the neck axis and *y*-direction is vertical rotation around the axis joining the ears (Fig. 4). Three additional gyroscope signals are derived from the *x*-direction and *y*-direction gyroscope signals; absolute angular velocity, a 10-point Moving Average Filter applied to the absolute angular velocity signal, and a 50-point Moving Average Filter applied to the absolute angular velocity signal. The absolute angular velocity $\alpha(n)$ is defined in Eq. (1) as the square-root of the sum of the squared *x*- and *y*-direction gyroscope signals. This value quantifies the angular velocity that takes place in any direction, giving a measure of overall head-movement and indicating whether or not head-movement has occurred.

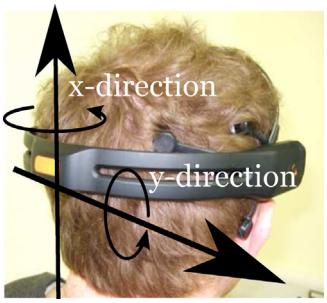

Fig. 4. Emotiv EPOC headset: gyroscope axes.

$$\alpha(n) = \sqrt{gyro_x(n)^2 + gyro_y(n)^2} \quad (1)$$

A 10-point MA filter was applied to the absolute angular velocity signal $\alpha(n)$ as described in Eq. (2).

$$MA_{10} = \frac{1}{10}\sum_{k=0}^{9}\alpha(n-k). \quad (2)$$

Similarly, Eq. (3) shows how the 50-point MA filter was applied to $\alpha(n)$:

$$MA_{50} = \frac{1}{50}\sum_{k=0}^{49}\alpha(n-k). \quad (3)$$

As was the case for the EEG signal, gyroscope data from each channel was separately pre-processed and segmented into overlapping windows. A set of 16 features (Table 2) was extracted from each of the five gyroscope channels, providing a total of 80 gyroscope features extracted for each window. A detailed description of the gyroscope features is included in Appendix A. An overview of the gyroscope classifier architecture is displayed in Fig. 5.

Support vector machines classifier. In each of the classifiers described in this system, the extracted feature vectors are classified using a support vector machines (SVMs) classifier. The fundamental concept of the support vector machine is to transform the data into a higher dimensional space and to find the optimal hyperplane in that space to maximise the margin between classes. By transforming the data into a higher dimensional space, complex classification problems can thus be simplified, allowing the use of linear discriminant functions to effectively separate the data (Cristianini and Shawe-Taylor, 2000). As the SVM is based only on those training patterns that are near the decision surface, classification calculation times are usually small, allowing for ease of implementation in a real-world, ambulatory system.

Classification consists of two steps: training and testing. Leave-one-out cross validation (LOO-CV) was used to assess the

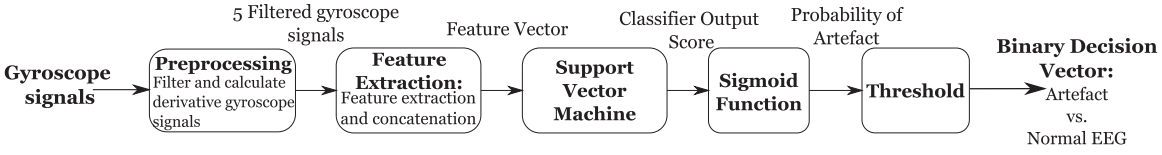

Fig. 5. System architecture for gyroscope artefact detector.

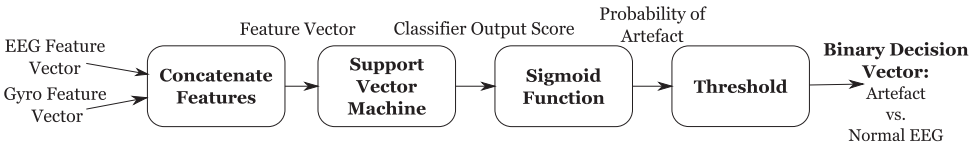

Fig. 6. Overview of the feature fusion classifier combination architecture. Feature fusion as illustrated in the diagram is repeated for each EEG channel.

performance of the system for participant independent head-movement artefact detection. The classifier training data was thus sourced from 6 participants and tested on the remaining participant. This procedure was repeated until each participant had been a test subject and the mean result is reported. The leave-one-out method is known to be an almost unbiased estimation of the true generalisation error (i.e. the performance reported with the leave-one-out method is the most similar to the performance that this system would show on an unseen test dataset of infinite length once it is trained on all available data, Vapnik, 2006).

In the training stage, the training data was normalised by subtracting the mean and dividing by the standard deviation to ensure that all features have equal significance when training the model. Gaussian kernel SVM classifiers have proven effective in a number of other event detection systems in EEG (Kelleher et al., 2010; Temko et al., 2011). A Gaussian radial basis function kernel SVM was implemented in this study (Scholkopf and Smola, 2002). For model selection (i.e. to search for the optimal Gaussian kernel parameter and generalisation parameter), fivefold cross validation on the training data was used. Once found, these parameters are used to train the final SVM model on all the training data.

In the testing stage, the normalisation template from the training stage was applied to the testing data. The trained classifier was then applied to each channel of the normalised testing data separately, and the decisions were post-processed and fused as described below.

Post-processing. The output from the SVM was obtained for each window in every channel. This output was converted to a posterior probability by approximating the posterior with a sigmoid function:

$$P(Art|x) \approx P_{A,B}(f) = \frac{1}{1+\exp(Af+B)} \tag{4}$$

where $f = f(x)$ is the distance to the separating hyperplane (i.e. the output of the SVM classifier), and A and B are the parameters of the sigmoid function estimated on the training dataset (Platt, 1999). This probability output $P(Art|x)$ was compared to a threshold value θ, resulting in a binary output decision; 1 indicating artefact and 0 indicating normal EEG.

2.3. Classifier combinations

The objective of the classifier combinations outlined in this paper is to take advantage of information regarding the presence of artefacts from both available sets of signals (EEG and gyroscope). There are three methods of classifier combination investigated: feature fusion, decision fusion and score fusion.

2.3.1. Feature fusion

In the feature fusion classifier (Fig. 6), EEG and gyroscope features are extracted as per the individual EEG and gyroscope systems. For each window, these features are concatenated to produce a single SVM model. Thus, each channel of EEG has 69 EEG and 80 gyroscope features associated with it for each window. This allows information from the gyroscope signal to be incorporated into the classification task for per-channel artefact detection. As was the case for the individual EEG and gyroscope classifiers, a probability of the artefact-presence is obtained for each epoch, and a binary decision vector is subsequently produced using the sigmoid function.

2.3.2. Score fusion

In the score fusion method of classifier combination, EEG and gyroscope classifiers are trained and applied separately to the data and combined afterwards. The gyroscope classifier output probabilities are combined with the output probabilities from each EEG channel to allow per-channel artefact detection. In this paper, score fusion methods using non-trainable, fixed combination rules are used. These fixed rule combiners are ready for combination as soon as the base classifiers are trained and make use of the fact that the outputs of the EEG and gyroscope classifiers have a clear interpretation, namely, they are the class posterior probabilities. For this two-class problem, the combined classifier $Q(\mathbf{x})$ is then:

$$Q(\mathbf{x}) = \mathcal{F}[C(\mathbf{x})] \tag{5}$$

where $Q(\mathbf{x}) = Q_j(\mathbf{x})$, $j = 1, \ldots, c$ for j base classifiers and $\mathcal{F}$ is a fixed-rule combination function. Several fixed-rule, score-level combination functions are investigated in this work: mean, product, maximum and minimum rules. An overview of the score fusion classifier architecture is illustrated in Fig. 7.

Mean/sum rule. The mean rule is simply the average of the classifier posterior probabilities: $Q(\mathbf{x}) = (1/L)\sum_j^L C_j(\mathbf{x})$, for j classifiers. In combining the EEG and gyroscope classifiers, $Q(\mathbf{x})$ is thus the average of the EEG and gyroscope classifier posterior probabilities. In some papers, the scaling factor L is omitted, and the combination output $Q(\mathbf{x})$ simply becomes the sum of the posterior probabilities.

Product rule. The product rule gives a combined output probability as the product of the posterior probabilities of the base classifiers

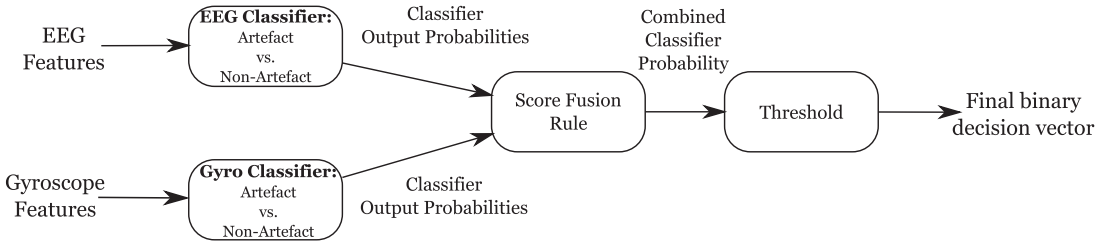

Fig. 7. Overview of the score fusion classifier combination architecture. The process illustrated in the diagram is repeated for each EEG channel.

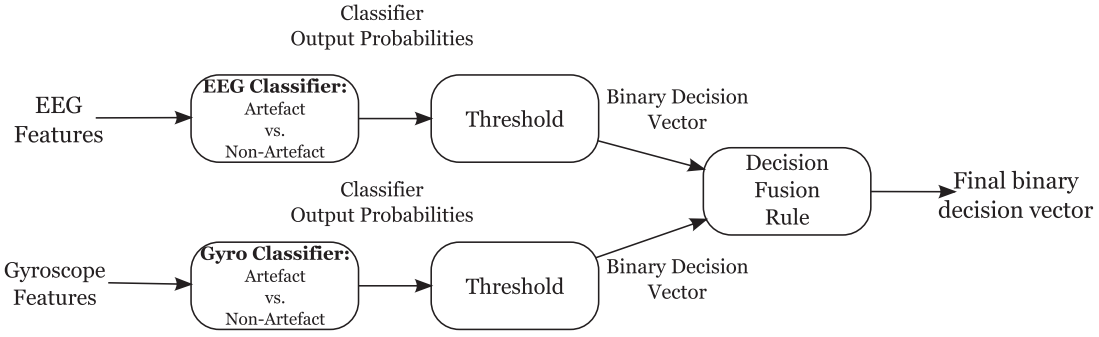

Fig. 8. Overview of the decision fusion classifier combination architecture. The decision fusion process highlighted in the diagram is repeated for each channel of EEG.

such that, $Q(\mathbf{x}) = \prod_j C_j(\mathbf{x})$. The product rule has been shown to be an effective method of combining classifiers at the score-level if the individual classifiers are independent, e.g. if the classifiers have been created for different feature spaces that are entirely unrelated. The product rule assumes that the confidence estimates are reliable and noise-free (Tax et al., 2000). It fails if these confidence estimates are very small or zero. The product rule for combining the EEG and gyroscope classifiers is thus given by product of the EEG and gyroscope classifier posterior probabilities.

Maximum rule. The maximum rule selects the classifier that is most confident of the classification decision, i.e. $Q(\mathbf{x}) = \max_j\{C_j(\mathbf{x})\}$. However, the maximum rule can fail if one of the classifiers is overtrained; the overtrained classifier is then overconfident of its decision, dominating the outcome, without improving the classification performance. The maximum rule also underperforms if there are simple component classifiers that are insensitive to nuances in the classification task that more complicated classifiers are able to detect. The insensitive classifiers then tend to dominate the maximum rule, reducing performance (Tax et al., 2000). The maximum rule for combining the EEG and gyro classifiers is thus the higher of the two posterior probabilities.

Minimum rule. The minimum rule selects the outcome of the classifier that has the least objection against a certain class, i.e. the minimum of the EEG and gyroscope classifier posterior probabilities such that $Q(\mathbf{x}) = \min_j\{C_j(\mathbf{x})\}$.

2.3.3. Decision fusion

As was the case for the score fusion methods, the EEG and gyroscope classifiers were trained and tested separately and combined on a per-channel basis afterwards. Decision level classifier fusion is accomplished by combining the binary output decisions $\mathcal{D}_j(\mathbf{x})$ from the individual classifiers (after the threshold θ has been applied to the classifier output probabilities), where the binary decisions are represented by Eqs. (6) and (7). In the decision fusion methods investigated in this paper, the same threshold was applied to the EEG classifier and to the gyroscope classifier, such that $\theta_{\text{EEG}} = \theta_{\text{Gyro}}$. Two decision fusion methods are investigated: logical OR and logical AND.

$$D_{\text{EEG}}(\mathbf{x}) = \begin{cases} 0 & \text{EEG classifier decides normal EEG} \\ 1 & \text{EEG classifier decides artefact} \end{cases} \quad (6)$$

$$D_{\text{Gyro}}(\mathbf{x}) = \begin{cases} 0 & \text{Gyroscope classifier decides normal EEG} \\ 1 & \text{Gyroscope classifier decides artefact} \end{cases} \quad (7)$$

Logical AND method. This classifier fusion method is performed using the logical AND operation, with 1 indicating the presence of artefact. The combined classifier output $\mathcal{D}(\mathbf{x})$ is then given by $\mathcal{D}(\mathbf{x}) = \mathcal{D}_{\text{EEG}}(\mathbf{x})\text{AND}\mathcal{D}_{\text{Gyro}}(\mathbf{x})$. If 1 exists in both classifier decisions, then the result of the combination of the two decision vectors will also be a 1, indicating that artefact is present in the epoch in question. Thus, if an artefact is to be detected using the logical AND method, both classifiers must detect the presence of the artefact individually.

Logical OR method. This decision fusion technique examines the combination of EEG and gyroscope classifier binary output decisions using the logical OR operator. The combined classifier output $\mathcal{D}(\mathbf{x})$ is then given by $\mathcal{D}(\mathbf{x}) = \mathcal{D}_{\text{EEG}}(\mathbf{x})\text{OR}\mathcal{D}_{\text{Gyro}}(\mathbf{x})$. If the binary decision vector from either EEG or gyroscope classifier detects artefact, the result of the logical OR classifier combination will report artefact. This process is illustrated in Fig. 8.

2.4. Classifier performance metrics

Epoch based metrics are used to evaluate classifier performance. In this binary decision problem (artefact vs. non-artefact), the decision made by the classifier can be represented in a structure known as a contingency table. This contingency table has four categories: true positives (*TP*) are epochs correctly labeled as artefact; false positives (*FP*) refer to epochs incorrectly labeled as artefact; true negatives (*TN*) correspond to correctly labeled non-artefact epochs and false negatives (*FN*) refer to epochs incorrectly labeled as non-artefact. The accuracy of each class is then evaluated using *Sensitivity* and *Specificity*. *Sensitivity* relates to the classifiers ability to correctly identify positive results (in this case, artefact) and is defined as:

$$\textit{Sensitivity} = \frac{TP}{TP + FN}. \quad (8)$$

Specificity gives a measure of a classifiers ability to identify negative results (in this experiment, normal EEG) and is defined as:

$$\textit{Specificity} = \frac{TN}{TN + FP}. \quad (9)$$

Receiver operator characteristic (ROC) curves show how sensitivity varies with specificity (Fawcett, 2006). The area under the ROC curve is an effective way of comparing the performance of two different systems. A random discrimination will give an area of 0.5 under the curve, while perfect discrimination will give unity area under the ROC curve. In this paper, the area under the ROC curve (ROC area) is calculated using trapezoidal numerical integration.

As there often exists a large skew in the class distributions of EEG event detection classifiers, the use of precision-recall (PR) curves in conjunction with ROC curves is often advocated so as to accurately evaluate the expected performance of a classification algorithm applied to unseen data. In the artefact detection experiment outlined in this paper, the class distributions are largely

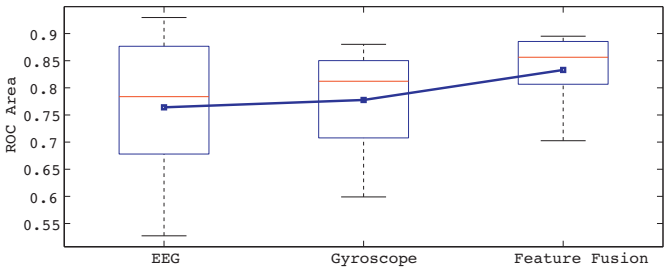

Fig. 9. Boxplot of ROC areas for feature fusion classifier with epoch length of 1 s. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

equal making the use of PR curve redundant (Davis and Goadrich, 2006).

3. Results and discussion

All classifier fusion methods investigated in this paper showed improved performance in the head-movement artefact detection task when compared to the individual EEG and gyroscope classifiers.

Fig. 9 displays boxplots of the ROC areas across participants for the EEG, gyroscope and feature fusion classifiers. Both mean and median ROC areas of the feature fusion classifier are higher than those of the EEG and gyroscope classifiers. This indicates that the EEG and gyroscope features contain complementary information in discriminating between head-movement artefact and background EEG, and consequently a classifier that employs features from both modalities will outperform both individual EEG and gyroscope classifiers. This net gain in classifier performance is accompanied by a reduction in the variability of ROC areas across participants for the feature fusion classifier, offering more robust classification in the face of changing users.

Fig. 10 illustrates that combining the individual EEG and gyroscope classifiers using any of the fixed-rule score fusion combinations results in improved performance in the head-movement artefact detection task. For each of the score fusion combination rules, the mean ROC area is higher than those of either the EEG or gyroscope classifiers alone. Similarly, the median ROC areas for each of the score fusion classifiers is higher than that of the EEG classifier and equal to or greater than that of the gyroscope classifier. As was the case for the feature fusion classifier, each of the score fusion combination rules, provides a lower variability in ROC areas that adds a level of robustness to the head-movement classification task. Amongst the score fusion combination rules, the mean/sum rule provides the greatest improvement in classifier performance. This is further highlighted in Fig. 11, which displays the mean ROC plots for each of the investigated score fusion combination rules

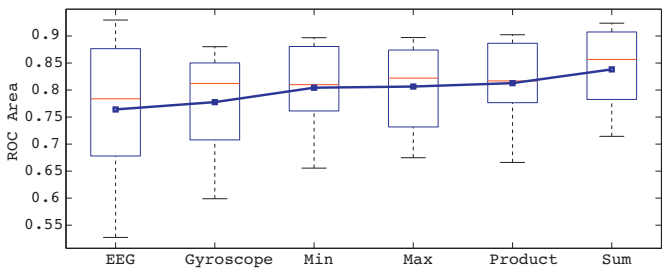

Fig. 10. Boxplot of ROC areas for score fusion classifiers with epoch length of 1 s. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are plotted in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

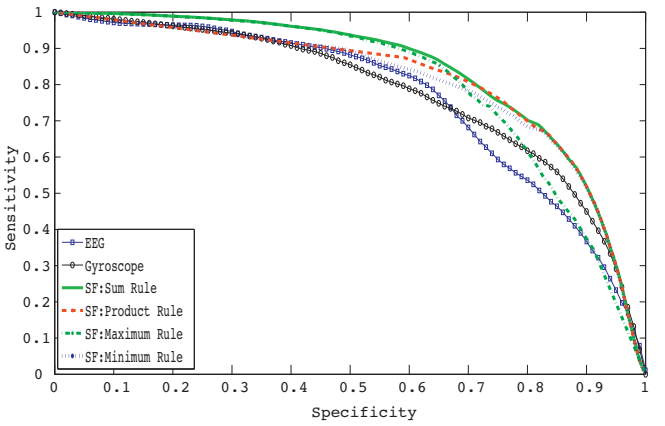

Fig. 11. Mean ROC plots for score fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed.

as well as those of the EEG and gyroscope classifiers, where the sum rule can be seen to dominate (i.e. lie above and to the right of) each of the other classifiers. It is also evident that the performance of the other combination rules matches that of the sum rule over defined sections of the ROC plot. For example, in applications where head-movements must be detected with a high degree of accuracy (i.e. high specificity), deploying the product or minimum rule in place of the sum rule will suffice. Similarly, for applications where detecting as many artefacts as possible is of concern (i.e. high sensitivity), utilising the maximum rule will provide the same performance as that of the sum rule. However, as there exists little or no difference in computational complexity between the rules, if score-level fusion of classifiers is to be employed then the sum rule should be used. This result is consistent with findings from other classifier combination tasks. Kittler et al. (1998) found that for an identity verification task, that amongst score fusion combination rules, the sum rule is most resilient to estimation errors. Tax et al. (2000) showed that in a two-class problem in which the posterior probabilities are well estimated and with completely independent feature spaces, the product rule performs best. However, when these assumptions are violated, the sum rule outperforms the product rule. In the head-movement artefact classification task, it may be unreasonable to assume that the EEG and gyroscope signals are independent. Although the feature spaces are derived from different physiological signals, the class labels for both modalities are dependent on the EEG signal.

Boxplots of the ROC areas across participants for the decision fusion classifiers are presented in Fig. 12. Both logical OR and logical AND fusion methods outperform the EEG classifier as measured by mean and median ROC areas. Mean ROC plots for the logical OR decision fusion method and the logical AND combination are displayed in Fig. 13. As expected, the logical OR combination rule is

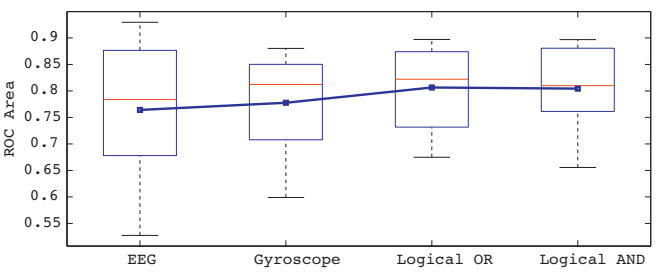

Fig. 12. Boxplot of ROC areas for decision fusion classifiers with epoch length of 1 s. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

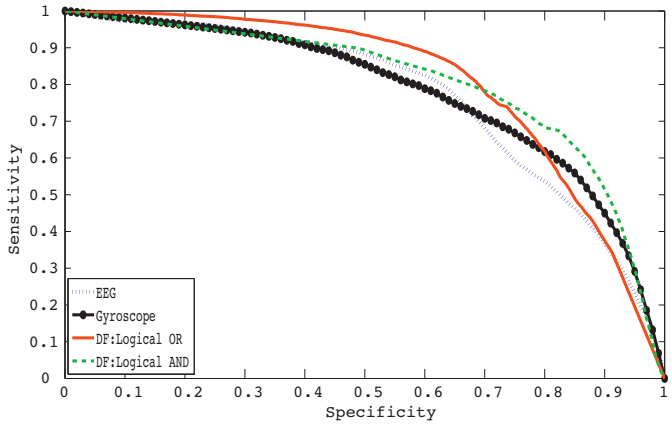

Fig. 13. Mean ROC plots for decision fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed.

highly sensitive to the detection of artefacts; if either EEG or gyroscope classifier detects a head-movement artefact, the logical OR output will flag that EEG epoch as artefact. The logical OR combination rule thus outperforms the logical AND classifier and the individual EEG and gyroscope classifiers for specificities below 0.65. However, this increased sensitivity comes at a significant reduction in performance for high specificities, where the logical OR rule falls below those of the individual EEG and gyroscope classifiers. In contrast to this, the logical AND combination rule is highly conservative and only marks a section of EEG as artefact if both EEG and gyroscope classifiers agree, and consequently performs best if high specificities are required. In this regard, the logical OR and the logical AND are analogous to the score fusion maximum and minimum rules, respectively.

The feature fusion and top-performing score and decision fusion classifiers are compared in Fig. 14. As the decision-level fusion offered similar classification performance to the minimum and maximum score fusion combination rules, it is unsurprising that the sum rule outperforms the best decision fusion classifier. The mean and median ROC areas of the feature fusion and sum rule score fusion are approximately equal. However, the variability in ROC areas of the sum rule is larger than that of the feature fusion. Accordingly, for the head-movement classification task where EEG and gyroscope features are available, combination of these signals at the feature level is advised. However, for classifiers other than support vector machines, score fusion may be more appropriate. SVMs do not suffer from the "curse of dimensionality" to the same extent as many other classifier models; hence, combining the different modalities at the feature level is unlikely to lead to excessive overfitting. For classifiers that are highly susceptible to overfitting, such as neural networks, feature fusion may do just this. It is proposed that this may indeed

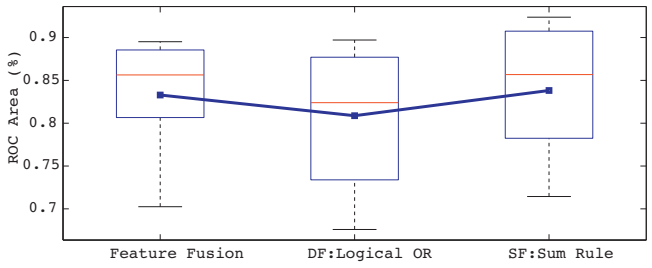

Fig. 14. Boxplot of ROC areas for the best performing feature, score and decision fusion classifiers. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

be the case in the systems outlined by Malarvili and Mesbah (2008) and Bermudez et al. (2007), where neural networks were deployed and classifier fusion was advocated in place of feature fusion.

As the performance of feature fusion and score fusion (sum rule) was so similar, there exist a number of scenarios whereby use of the score fusion sum rule makes more sense. Score fusion offers the advantage that the artefact detection system will continue to operate if the gyroscope signal becomes unavailable. In our experience of collecting clinical data, it is often the case that additional physiological signals are not recorded or partially recorded, particularly in longer recordings. In such an eventuality, the classifier combination stage outlined in this paper can be switched off, and the EEG artefact detection classifier can continue alone. Where the gyroscope to become unavailable in the feature fusion case, the system would not be able to operate, as the gyroscope features would now lie in an unexpected region of the feature space, leading to likely adverse classification performance. Of course this feature requires the recognition by the clinician/researcher that the gyroscope signal is unavailable or corrupted. This step would be automated if the maximum score fusion or the logical OR decision fusion rules were adopted, albeit at a loss in classification accuracy compared to feature fusion or score fusion, sum rule.

Although outside the scope of this paper, score fusion offers a further advantage in that classifiers from additional modalities can easily be incorporated to the artefact detection system, without re-training of the classifiers. Thus, if additional physical signals such as accelerometer data or electromyogram signals were to be made available, score fusion combination could easily accommodate these. For marginally better classifier performance, feature fusion would require the classifier to be entirely re-trained incorporating features from the added physiological signals.

The multimodal head-movement artefact detection system presented in this paper is suitable for effective artefact annotation of ambulatory EEG recordings. Discussions with clinicians and researchers at Cork University Hospital and University College London Hospitals indicate that this would be a desirable resource in helping clinicians and researchers to disambiguate EEG. Alternatively, in automated neurological event detection systems, the artefact detection system outlined in this paper can be directly implemented as a preprocessing stage whereby epochs that contain EEG artefacts are rejected before being sent to the neurological event detection classifier. This process is common in many event-related potential experiments (Luck, 2005) and for some BCI applications (Fatourechi et al., 2007). However, in certain neurological event detection systems, it can be highly undesirable to discard large sections of EEG, most notably when the loss of data pertaining to infrequent events is unacceptable, such as in interictal epileptiform activity detection. In this scenario, head-movement artefact detection could be incorporated at the post-processing stage by an additional layer of decision fusion between the artefact detector and the neurological event detector. Alternatively, the artefact detection system outlined here could be incorporated in a 3-class classification set-up with a classifier discriminating between background EEG, artefact and the neurological event under question. A final alternative for including the multimodal artefact detection system described in this paper in an automated neurological event detection system would be to incorporate it into an automated blind source separation artefact removal stage (such as independent component analysis and canonical correlation analysis), where estimated EEG sources components are retained or removed based on the output of the multimodal artefact detection classifier. Each of these alternatives requires the collection of suitable data, and in the case of blind source separation, annotation of the estimated source signals and will form the basis of future work.

4. Conclusion

In light of examining a range of multimodal classifier fusion methods incorporating both EEG and gyroscope signals, it was found that the fusion of signals at both feature and classifier levels improved the detection of head-movement artefact in ambulatory EEG compared to using either EEG or gyroscopes alone, confirming the complimentary nature of information carried by these different modalities. This result was repeated for each of score-level and decision-level fixed-rule combinations that were investigated. Amongst the methods investigated, it was found that feature fusion and the score-level, sum rule offered the best classification performance by increasing the ROC areas and reducing inter-participant variability. By illustrating the feasibility of fusing EEG and gyroscope signals at the classifier level, a number of implementation options are unlocked; most notably, robustness in the face of intermittent gyroscope signals and easy integration of additional physiological signals into the artefact detection task, should they become available.

Conflict of interest statement

Simon O'Regan and William Marnane declare no conflict of interest.

Acknowledgement

This research has been possible thanks to Science Foundation Ireland grants SFI/07/SRC/I1169 and SFI/10/IN.1/B3036.

Appendix A. Gyroscope features

Gyroscope features were chosen to reflect the fact that when head-movement occurs, the gyroscope signal undergoes relatively large deflections that oscillate with changing direction.

A.1. No. of zero crossings of the gyroscope signal and its derivatives

The number of zero crossings (Zc) is the number of times within an epoch that the gyroscope signal crosses the *x*-axis. The number of zero crossings of the 1st derivative of the gyroscope corresponds to the number of local maxima and minima of the gyroscope signal. The number of zero crossings of 2nd derivative corresponds to the number of times that the 2nd derivative of the EEG signal crosses the *x*-axis within an epoch. During periods when the participant remains still, the gyroscope signal is centered about the *x*-axis with small deviations representing minute movements of the headset and consequently crosses the *x*-axis frequently. Due to the large deflections in the gyroscope signals when head-movement occurs, the number of zero crossings of the gyroscope and its derivatives should be smaller in sections of head-movement.

A.2. Nonlinear energy (NLE)

NLE was introduced by D'Alessandro et al. (2003) as a feature in an epileptic seizure prediction algorithm. Non-linear energy (NLE) is a function of the amplitude of a signal, and the change of that amplitude and is defined as:

$$\mathrm{NLE}(\mathbf{x}_j) = \frac{1}{n_s - 2}\sum_{i=2}^{n_s-1} x_j(i)^2 - x_j(i-1)x_j(i+1). \tag{A.1}$$

Gyroscope signals during periods of head-movement typically contain more energy than periods without movement, and consequently should exhibit higher NLE.

A.3. Line length/curve length

Line length was introduced by Esteller et al. (2001) in the context of seizure detection. Line length (sometimes referred to as curve length) is defined for an epoch $\mathbf{x}_j$ as:

$$L(\mathbf{x}_j) = \sum_{i=0}^{n_s} |x_j(i+1) - x_j(i)|. \tag{A.2}$$

Thus, line length is the running sum of distances between consecutive points within the sliding window of size n_s. Accordingly, gyroscope signals corresponding to movements, where there are big amplitude jumps between samples, are expected to have considerably higher line length than gyroscope signals corresponding to no head-movements.

A.4. Maximum value of autocorrelation

The maximum value of the autocorrelation $R_{ss}(\mathbf{x}_j)$ of a gyroscope signal was used by Tunçel et al. (2009) to classify human leg movements. The autocorrelation is defined as:

$$R_{ss}(\mathbf{x}_j) = \frac{1}{N_s - \Delta}\sum_{i=0}^{N_s-\Delta-1} (s_i - \mu_s)(s_{i-\Delta} - \mu_s), \tag{A.3}$$

where $\Delta = 0, 1, \ldots, N_s - 1$. Repetitive movements such as those produced during shaking of the head have higher maximum autocorrelation values than irregular movements or periods of rest.

A.5. No. of inactive samples

The number of inactive samples within an epoch $\mathbf{x}_j$ is defined as the number of samples for which there is very little change in the gyroscope amplitude. This was calculated by applying a threshold of 0.01 to the absolute value of the derivative of the gyroscope signal. Gyroscope sections without head-movements should therefore have higher numbers of inactive samples.

A.6. Activity/variance (1st Hjorth parameter)

Hjorth (1970) introduced a number of features to characterise seizure EEG that have since become widespread in their use in the context of EEG signal processing. These features are utilised here for gyroscope signals. The 1st Hjorth parameter is simply the variance of the signal and given by:

$$\mathrm{Activity}(\mathbf{x}_j) = \sigma_j^2 = \frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j)^2. \tag{A.4}$$

Owing to the fact that a large range of head-movements are examined in this experiment, it is anticipated that the variance of gyroscope signals corresponding to artefactual sections of EEG should on average be greater than the variance of those corresponding to normal EEG.

A.7. Mobility (2nd Hjorth parameter)

The Hjorth mobility of an epoch is defined as:

$$\mathrm{Mobility}(\mathbf{x}_j) = \frac{\sigma_{\Delta j}}{\sigma_j}, \tag{A.5}$$

where $\sigma_{\Delta j}$ is the standard deviation of the first derivative of an epoch.

A.8. Complexity (3rd Hjorth parameter)

The complexity of an epoch is defined as:

$$\text{Complexity}(\mathbf{x}_j) = \frac{\sigma_{\Delta 2j}/\sigma_{\Delta j}}{\sigma_{\Delta j}/\sigma_j}, \tag{A.6}$$

where $\sigma_{\Delta 2j}$ is the standard deviation of the second derivative of the epoch. Gyroscope signals that correspond to head-movements tend to produce brief undulations and should therefore exhibit less complexity (due to more regular gyroscope signals) than those of gyroscope epochs corresponding to normal EEG.

A.9. RMS amplitude

The root mean square (RMS) amplitude, or quadratic mean, is a statistical measure of the magnitude of a time varying quantity. The RMS amplitude expresses the mean of the absolute amplitude of an epoch $\mathbf{x}_j$ and is defined as:

$$\text{RMS}(\mathbf{x}_j) = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s} x_j^2(i)}. \tag{A.7}$$

As gyroscope signals corresponding to artefact EEG are often high-energy, high-amplitude signals, the RMS amplitude aims to capture this trait.

A.10. Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. Skewness has been used to classify human leg movements with gyroscopes in Tunçel et al. (2009) and is defined as:

$$\text{Skewness}(\mathbf{x}_j) = \frac{(1/n_s)\sum_{i=1}^{n_s}(x_j(i)-\mu_j)^3}{((1/n_s)\sum_{i=1}^{n_s}(x_j(i)-\mu_j)^2)^{3/2}}, \tag{A.8}$$

where $\mu_j(\mathbf{x}_j)$ is the mean of an epoch $\mathbf{x}_j$ and defined as:

$$\mu_j(\mathbf{x}_j) = \frac{1}{n_s}\sum_{i=1}^{n_s} x_j(i). \tag{A.9}$$

Owing to the variety of different head-movements, the distribution of gyroscope signals corresponding to sections of EEG where artefacts are produced is hypothesised to have higher skew than those corresponding to the condition of no movement.

A.11. Kurtosis

Kurtosis, often referred to as the fourth central moment, is a measure of the "peakedness" of a probability density function and is defined as follows:

$$\text{Kurtosis}(\mathbf{x}_j) = \frac{(1/n_s)\sum_{i=1}^{n_s}(x_j(i)-\mu_j)^4}{((1/n_s)\sum_{i=1}^{n_s}(x_j(i)-\mu_j)^2)^2}, \tag{A.10}$$

and has been used to classify human leg movements by Tunçel et al. (2009).

A.12. Variance of first and second derivatives

The variance of the first ($\sigma_{\Delta j}$) and second ($\sigma_{\Delta 2j}$) derivatives was used by Thomas et al. (2010) for EEG seizure detection. They are included here as features in the artefact classification task as they were found to provide reasonable separation between gyroscope signals during rest and signals during movements.

A.13. Total power

The total power refers to the sum of power in all bins of the power spectral density (PSD) between 0 and 12 Hz:

$$P_{\text{total}}(\mathrm{x_j}) = \sum_{i=0}^{12n_s/fs} a_j(i), \tag{A.11}$$

where $a_j(i)$ is the power in bin i of epoch $\mathbf{x}_j$. This feature is common in gyroscope classifiers and refers to the total power in the frequency range of typically occurring gyroscope signals. Periods of movement usually exhibit higher total power than inactive sections.

References

Allison B, Wolpaw E, Wolpaw J. Brain–computer interface systems: progress and prospects. Expert Rev Med Devices 2007;4(4):463–74.

Anderson E, Potter K, Matzen L, Shepherd J, Preston G, Silva C. A user study of visualization effectiveness using EEG and cognitive load. Comput. Graphics Forum, vol. 30. Wiley Online Library; 2011. p. 791–800.

Barlow J. Automatic elimination of electrode-pop artifacts in EEG's. IEEE Trans Biomed Eng 1986;33(5):517–21.

Berg P, Scherg M. Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG. Clin Phys Physiol Meas 1991;12(A):49–54.

Bermudez T, Lowe D, Lamborelle A. Schemes for fusion of EEG and ECG towards temporal lobe epilepsy diagnostics. In: Proc IEEE Eng Med Biol Soc (EMBC); 2007. p. 5132–5.

Bobrov P, Frolov A, Cantor C, Fedulova I, Bakhnyan M, Zhavoronkov A. Brain–computer interface based on generation of visual images. PLoS ONE 2011;6(6):e20674.

Brunner D, Vasko R, Detka C, Monahan J, Reynolds C III, Kupfer D. Muscle artifacts in the sleep EEG: automated detection and effect on all-night EEG power spectra. J Sleep Res 1996;5(3):155–64.

Casson A, Rodriguez-Villegas E. Utilising noise to improve an interictal spike detector. J Neurosci Methods 2011;201(1):262–8.

Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press; 2000.

Croft RJ, Barry RJ. Removal of ocular artifact from the EEG: a review. Clin Neurophysiol 2000;30(1):5–19.

D'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echauz J, Litt B. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. IEEE Trans Biomed Eng 2003;50(5):603–15.

Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Proc Int Conf Mach Learn (ICML); 2006. p. 233–40.

Debener S, Minow F, Emkes R, Gandras K, Vos M. How about taking a low-cost, small, and wireless EEG for a walk? Psychophysiology 2012;49(11):1617–21.

Delorme A, Makeig S, Sejnowski T. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In: Proc Int ICA Conf; 2001. p. 457–62.

Duvinage M, Castermans T, Petieau M, Hoellinger T, De Saedeleer C, Seetharaman K, et al. A P300-based quantitative comparison between the Emotiv EPOC headset and a medical EEG device. In: Proc IASTED Int Conf Biomed Eng; 2012. p. 2012.

Emotiv EPOC headset. http://www.emotiv.com/ [accessed 27.09.12].

Esteller R, Echauz J, Tcheng T, Litt B, Pless B. Line length: an efficient feature for seizure onset detection. In: Proc IEEE Eng Med Biol Conf (EMBC); 2001. p. 1707–10.

Fatourechi M, Bashashati A, Ward R, Birch G, et al. EMG and EOG artifacts in brain computer interface systems: a survey. Clin Neurophysiol 2007;118(3):480–94.

Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27(8):861–74.

Gasser T, Schuller J, Gasser U. Correction of muscle artefacts in the EEG power spectrum. Clin Neurophysiol 2005;116(9):2044–50.

Gasser T, Ziegler P, Gattaz WF. The deleterious effect of ocular artefacts on the quantitative EEG, and a remedy. Eur Arch Psychiatry Clin Neurosci 1992;241(6):352–6.

Goncharova I, McFarland D, Vaughan T, Wolpaw J. EMG contamination of EEG: spectral and topographical characteristics. Clin Neurophysiol 2003;114(9):1580–93.

Greene B, Boylan G, Reilly R, de Chazal P, Connolly S. Combination of EEG and ECG for improved automatic neonatal seizure detection. Clin Neurophysiol 2007;118(6):1348–59.

Gwin J, Gramann K, Makeig S, Ferris D. Removal of movement artifact from high-density EEG recorded during walking and running. J Neurophysiol 2010;103(6):3526–34.

Hjorth B. EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 1970;29(3):306–10.

Kapoor A, Shenoy P, Tan D. Combining brain computer interfaces with vision for object categorization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2008. p. 1–8.

Kelleher D, Temko A, Nash D, McNamara B, Marnane W. SVM detection of epileptiform activity in routine EEG. In: Proc IEEE Eng Med Biol Conf (EMBC); 2010. p. 6369–72.

Khushaba R, Wise C, Kodagoda S, Louviere J, Kahn B, Townsend C. Consumer neuroscience: assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking. Expert Syst Appl 2013;40(9):3803–12.
Kittler J, Hatef M, Duin R, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20(3):226–39.
Lan T, Erdogmus D, Adami A, Mathan S, Pavel M. Channel selection and feature projection for cognitive load estimation using ambulatory EEG. Comput Intell Neurosci 2007;2007:8.
Lehmann C, Koenig T, Jelic V, Prichep L, John R, Wahlund L, et al. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). J Neurosci Methods 2007;161(2):342–50.
Lins O, Picton T, Berg P, Scherg M. Ocular artifacts in EEG and event-related potentials. I: Scalp topography. Brain Topogr 1993;6(1):51–63.
Lotte F, Fujisawa J, Touyama H, Ito R, Hirose M, Lécuyer A. Towards ambulatory brain-computer interfaces: a pilot study with P300 signals. In: Proc Int Conf Adv Comput Ent Technol; 2009. p. 336–9.
Luck S. An introduction to the event-related potential technique. MIT Press; 2005.
Malarvili M, Mesbah M. Combining newborn EEG and HRV information for automatic seizure detection. In: Proc IEEE Eng Med Biol Conf (EMBC); 2008. p. 4756–9.
Müller K, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring. J Neurosci Methods 2008;167(1):82–90.
O'Regan S, Faul S, Marnane W. Automatic detection of EEG artefacts arising from head movements. In: Proc IEEE Eng Med Biol Conf (EMBC); 2010a. p. 6353–6.
O'Regan S, Faul S, Marnane W. Automatic detection of EEG artefacts arising from head movements using gyroscopes. In: Proc Int Sympos Appl Sci Biomed Commun Technol (ISABEL); 2010b. p. 1–5.
O'Regan S, Faul S, Marnane W. Automatic detection of EEG artefacts arising from head movements using EEG and gyroscope signals. Med Eng Phys 2012., http://dx.doi.org/10.1016/j.medengphy.2012.08.017, in press.
Peng Y, Lin C, Sun M, Landis C. Multimodality sensor system for long-term sleep quality monitoring. IEEE Trans Bio-med Circuits Syst 2007;1(3):217–27.
Picton T, van Roon P, Armilio M, Berg P, Ille N, Scherg M. The correction of ocular artifacts: a topographic perspective. Clin Neurophysiol 2000;111(1):53–65.
Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Larg Margin Classifiers 1999;10(3):61–74.
Polikar R, Tilley C, Hillis B, Clark C. Multimodal EEG, MRI and PET data fusion for Alzheimer's disease diagnosis. In: Proc IEEE Eng Med Biol Conf (EMBC), vol. 20; 2010. p. 6058–61.
Qian M, Aguilar M, Zachery K, Privitera C, Klein S, Carney T, et al. Decision-level fusion of EEG and pupil features for single-trial visual detection analysis. IEEE Trans Biomed Eng 2009;56(7):1929–37.
Scholkopf B, Smola A. Learning with kernels. Cambridge, MA: MIT Press; 2002.
Tax D, Van Breukelen M, Duin R, Kittler J. Combining multiple classifiers by averaging or by multiplying? Pattern Recogn 2000;33(9):1475–85.
Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with support vector machines. Clin Neurophysiol 2011;122(3):464–73.
Thomas E, Temko A, Lightbody G, Marnane W, Boylan G. Gaussian mixture models for classification of neonatal seizures using EEG. Physiol Meas 2010;31(7):1047–64.
Thomas EM. A machine learning framework for neonatal seizure detection. Department of Electrical & Electronic Engineering, National University of Ireland; 2011, Ph.D.thesis.
Tunçel O, Altun K, Barshan B. Classifying human leg motions with uniaxial piezoelectric gyroscopes. Proc Soc Photo-opt Instrum 2009;9(11):8508–46.
Vapnik V. Estimation of dependences based on empirical data. Springer-Verlag New York Inc; 2006.
van de Velde M, Ghosh I, Cluitmans P. Context related artefact detection in prolonged EEG recordings. Comput Methods Prog Biomed 1999;60(3):183–96.
Waterhouse E. New horizons in ambulatory electroencephalography. IEEE Eng Med Biol Mag 2003;22(3):74–80.
Willis J, Nelson A, Rice J, Black F. The topography of muscle activity in quantitative EEG. Clin Electroencephalagr 1993;24(3):123–6.
Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav Brain Funct 2011;7(1):30.