

## Research Article

# Evaluation of Empirical Mode Decomposition for Event-Related Potential Analysis

**N. Williams, S. J. Nasuto, and J. D. Saddy**

*Centre for Integrative Neuroscience and Neurodynamics (CINN), University of Reading, Reading RG6 6AY, UK*

Correspondence should be addressed to N. Williams, sir07nw@reading.ac.uk

Received 2 July 2010; Revised 1 December 2010; Accepted 21 January 2011

Academic Editor: Patrick Flandrin

Copyright © 2011 N. Williams et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current methods for estimating event-related potentials (ERPs) assume stationarity of the signal. Empirical Mode Decomposition (EMD) is a data-driven decomposition technique that does not assume stationarity. We evaluated an EMD-based method for estimating the ERP. On simulated data, EMD substantially reduced background EEG while retaining the ERP. EMD-denoised single trials also estimated shape, amplitude, and latency of the ERP better than raw single trials. On experimental data, EMD-denoised trials revealed event-related differences between two conditions (condition A and B) more effectively than trials lowpass filtered at 40 Hz. EMD also revealed event-related differences on both condition A and condition B that were clearer and of longer duration than those revealed by low-pass filtering at 40 Hz. Thus, EMD-based denoising is a promising data-driven, nonstationary method for estimating ERPs and should be investigated further.

## 1. Introduction

Event-related potentials (ERPs) are changes in the ongoing electroencephalographic (EEG) signal related to processing of a stimulus [1]. The amplitude of ERP is usually much lower than that of the background EEG and is hard to visualise at the level of single trials. Because of this low signal-to-noise ratio, a number of single-trials time-locked to the presentation of the stimulus are averaged to obtain an estimate of the ERP. Single-subject averages are first obtained, and then, the single-subject averages are averaged across subjects to obtain a Grand Average. This procedure has a number of disadvantages. Firstly, when the signal-to-noise ratio (SNR) is particularly low, the number of single trials per participant can be high, leading to long experiment times and participant fatigue [2]. Secondly, the conventional averaging procedure is often unable to detect effects when the number of trials is low. An example of such a situation arises when examining differences between ERPs from trials when the subject responded correctly to when he/she responded incorrectly, and so forth. Thirdly, averaging does not reveal trial-to-trial variations in amplitude and/or latency of ERP. Thus, there is a need for methods which improve the

SNR of single trials thereby enabling subtle effects between conditions to be detected.

A number of methods have been proposed to perform such a filtering function. Model-based approaches using autoregressive models (AR) and autoregressive moving average models (ARMA) have been proposed [4, 5]. Approaches based on linear transforms such as Independent Component Analysis (ICA) and Principal Component Analysis (PCA) have also been applied [6–8] as have neurofuzzy techniques and methods based on the Wiener formalism [9, 10]. However, all these methods assume that the frequency content of the ERP is stationary with respect to time, while the ERP is recognised to actually be nonstationary in most cases [11].

A few time-varying denoising methods have also been proposed. de Weerd and Kap [12] introduced a time-varying Wiener filter while [11, 13] use the wavelet transform to obtain a multiresolution decomposition from which they each employ different criteria to select the event-related wavelet coefficients. Recently, [14] proposed a variant of this approach by applying wavelet decomposition to phase-space trajectories of raw single trials. While these methods do not assume stationarity, they do implicitly assume

knowledge of meaningful temporal scales in the data. This implicit knowledge is used to set the number of levels of decomposition for the wavelet analysis. Since the EEG is considered to contain energy at five different temporal scales corresponding to the gamma, beta, alpha, delta, and theta frequencies, the number of levels of wavelet decomposition is usually set to five.

However, this partitioning of temporal scales is somewhat arbitrary. Finer distinct temporal scales such as upper-alpha frequencies and lower-alpha frequencies have been found [15]. Hence, there is a need for a method that is data driven in that it decomposes the data according to time scales intrinsic to the data itself rather than making assumptions about meaningful temporal scales. In addition, the method should allow for the signal to be nonstationary.

Empirical Mode Decomposition (EMD) is a data driven method that has been used to analyse data from such diverse domains as meteorology and finance [16, 17]. It allows for the signal to be nonstationary. In biosignal analysis, it has been used for the analysis of EMG (electromyographic activity) in [18]. The possibility of denoising using partial reconstructions of the EMD-based decomposition was first proposed in [19]. In [20, 21] EMD-based denoising was shown to compare favourably against mean and median denoising and give better results than wavelet denoising for some signal classes. The evaluations were done on different classes of artificial signals and ECG (electrocardiographic) data. In the field of neuroscience, it has been used to estimate the ERP on high SNR Local Field Potentials (LFPs) data [22]. It has also been used in conjunction with the Hilbert Transform to detect transient periods of synchronisation in neuronal activity [23]. A multichannel extension of EMD has been applied to EEG responses from steady-state stimulation and shown to be able to estimate onset and offset of the steady-state response and also estimate the SSVEP (steady-state visual evoked potential) due to the stimulation [24, 25]. Recently, [26] demonstrated promising results from a procedure applying wavelet denoising to the individual modes from an EMD-based decomposition. The current paper presents results from a study evaluating the feasibility of using EMD for denoising event-related potentials in a low SNR environment.

Section 2 contains a description of the EMD technique. Section 3 presents details of how the simulated data were generated and how the experimental data were collected. Further, it outlines the analyses to be conducted on simulated and experimental data. Section 4 presents results of the analyses on simulated and experimental data. Section 5 discusses the results obtained and their implications, and Section 6 summarises the findings.

## 2. Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is a data driven method which extracts oscillatory modes intrinsic to the data by identifying time scales characteristic of individual oscillatory modes. It decomposes a signal into a small number of Intrinsic Mode Functions (IMFs) which represent the oscillatory modes contained in the data. The counterpart

of the IMF in Fourier analysis is the simple harmonic component. However, the IMF is much more general than the harmonic component since it can be modulated both in amplitude and frequency while a harmonic component has constant amplitude and frequency. The amplitude and frequency modulations are possible because the decomposition depends on the local characteristic time scale of the data. Because of this flexible nature of the decomposition, the IMFs are often physically meaningful. Also, because the decomposition depends on the local characteristic time scale, EMD is suitable for application to nonstationary signals. The working of EMD is described in detail in [3].

An IMF is identified by two criteria.

- (i) The number of extrema and number of zero-crossings must differ at most by one. This criterion eliminates riding waves and ensures a meaningful instantaneous frequency.
- (ii) At every discrete time point, the mean of the upper and lower envelopes should be zero. This ensures that the IMF is symmetric about the  $x$ -axis and eliminates fluctuations in the instantaneous frequency that might otherwise have been present.

IMFs are identified and extracted through a sifting process. The sifting process consists of a number of steps:

- (i) identify all extrema of the signal,  $X(t)$ ;
- (ii) interpolate between all maxima to get upper envelope;
- (iii) interpolate between all minima to get lower envelope;
- (iv) obtain the mean of the upper and lower envelopes,  $m_1$ .

The first component,  $h_1$ , should be

$$h_1 = X(t) - m_1. \quad (1)$$

Ideally,  $h_1$  should give the first IMF, but in practise, there is often overshoot or undershoot, and either or both criteria for IMF are not satisfied. When this happens,  $h_1$  is treated as the data such that

$$h_{11} = h_1 - m_{11}. \quad (2)$$

This is repeated  $k$  times until  $h_{1k}$  is an IMF or until the stopping criterion for the sifting process is reached

$$h_{1k} = h_{1(k-1)} - m_{1k}. \quad (3)$$

For the implementation of EMD that we used, the sifting process was stopped based on a double threshold criterion. The stopping criterion was formulated as a way of allowing for locally large excursions in the mean while ensuring globally small fluctuations. The criterion is described in detail in [27].

Then,  $h_{1k}$  is designated as the first IMF from the data:

$$c_1 = h_{1k}. \quad (4)$$

Overall,  $c_1$  should contain the highest frequency component of the signal.  $c_1$  is then subtracted from the original signal,  $X(t)$  to give the residual,  $r_1$ :

$$r_1 = X(t) - c_1. \quad (5)$$

Since  $r_1$  might contain further IMFs, it is subject to the same sifting process. This process can be iteratively applied to all subsequent residuals such that

$$r_n = r_{n-1} - c_n. \quad (6)$$

The extraction of IMFs stops when  $r_n$  becomes a monotonic function from which no more IMFs can be extracted. Thus, it follows that

$$X(t) = \sum_{j=1}^n c_j + r_n. \quad (7)$$

### 3. Methodology

**3.1. Simulated Data.** Tests were conducted on simulated data, to evaluate the denoising performance of EMD. Since the data were artificially generated, peak amplitude, peak latency, and morphology of ERP are known. Hence, denoising performance could be evaluated accurately.

**3.1.1. Data Generation.** For generating artificial ERP data, the additive model of ERP generation was assumed, which specifies that the ERP is additive to the background EEG [28]. Background EEG has previously been shown to be well modeled by an autoregressive (AR) process [29]. AR coefficients for a 3rd-order AR process were estimated by the Burg method, by fitting to a 1-second segment of background EEG obtained from the dataset in [30]. An order of 3 was chosen since it marked the steepest fall when forward prediction error was plotted against model order. The AR process was described by

$$x(t) = 1.6x(t-1) + 1.1x(t-2) + 0.4x(t-3) + r(t), \quad (8)$$

where  $x(t)$  was the AR-process and  $r(t)$  was the Gaussian white noise process driving the AR system.

The ERP was a smoothed version of a P300 obtained by averaging 48 target trials from a visual oddball experiment [30]. The ERP was scaled such that its peak amplitude was  $5 \mu\text{V}$ . Figure 1 displays the ERP over 1 second while Figure 2 shows a typical artificial raw single trial with the ERP added to the background EEG.

100 raw single trials were generated in this way with the standard deviation of the noise process adjusted to give an overall SNR of 1. SNR was calculated as the ratio of the variance of the ERP to the average variance of the 100 trials of background EEG.

**3.1.2. EMD Denoising.** Each artificial raw single-trial was decomposed using EMD. To obtain the EMD-denoised single trial, it is necessary to identify those IMFs carrying event-related energy and calculate their sum. Since the ERP is a low-frequency signal, we included all IMFs having dominant frequency below a threshold in the EMD-based single-trial reconstruction. We call this the threshold frequency ( $f_i$ ).

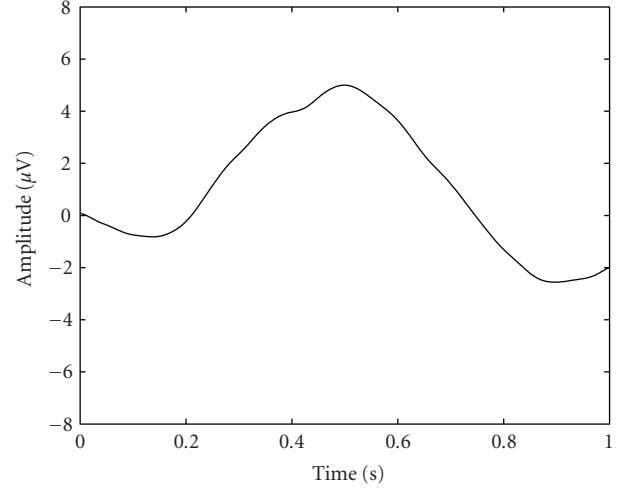


FIGURE 1: Scaled and smoothed version of an ERP called the P300 obtained by averaging 48 target trials from a visual oddball experiment [3].

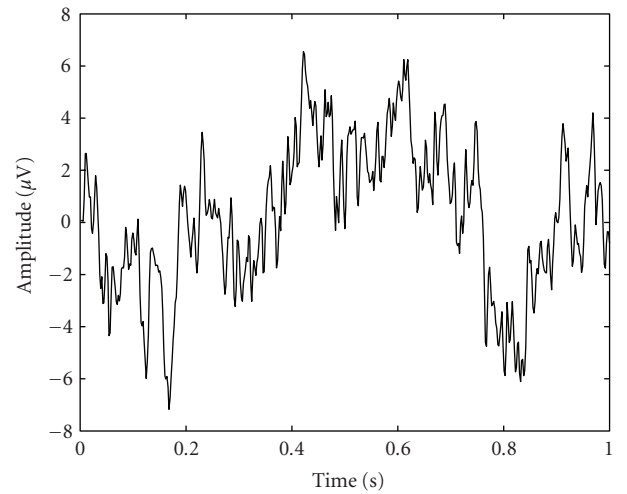


FIGURE 2: Typical artificial raw single-trial with ERP added onto background EEG. SNR = 1. The P300 ERP is hardly visible.

**Selecting Threshold Frequency ( $f_i$ ).** To develop a quantitative criterion for selecting a threshold frequency, we assumed a signal model

$$u(t) = s(t) + n(t), \quad (9)$$

where  $u(t)$  is a raw single trial and  $s(t)$  is the event-related potential (ERP) additive to the background EEG,  $n(t)$ . The performance of a denoising method has two aspects—the amount by which it reduces variance due to noise (background EEG) and the amount of variance due to signal (ERP) it retains.

A measure of noise reduction,  $N_f$ , can be obtained by calculating the reciprocal of average variance of the denoised single trials,  $x_{i=1,\dots,N}$  with higher values meaning better noise reduction. Assuming that the ERP is fully present in every

denoised single trial,  $N_f$  reflects the variance due to the background EEG removed by the denoising method

$$N_f = \frac{1}{1/N \sum_{i=1}^N \text{Var}(x_i)}. \quad (10)$$

A measure of signal retention,  $S_f$ , can be calculated as ratio of variance of denoised average,  $X_{f_D}$ , to variance of raw average,  $X_{f_R}$ , with higher values meaning better signal retention. Assuming that average of raw trials is the true ERP,  $S_f$  reflects the variance due to ERP retained in a denoised single trial

$$S_f = \frac{\text{Var}(X_{f_D})}{\text{Var}(X_{f_R})}. \quad (11)$$

Note that for a denoising method which does not discriminate between signal (ERP) and noise (background EEG),  $N_f$  and  $S_f$  are opposed to each other—high (low)  $N_f$  would imply a low (high)  $S_f$ . Thus, a measure of overall denoising performance,  $D_f$ , is given by combining  $N_f$  and  $S_f$  through multiplication

$$D_f = N_f S_f. \quad (12)$$

$D_f$  reflects the extent to which the ERP has been retained while also reducing the background EEG. It is a measure of the ability of the denoising method to discriminate between signal and noise.

This is illustrated by contrasting two situations—when the denoising method attenuates a large proportion of signal while retaining a large proportion of the noise,  $N_f$  and  $S_f$  will be low and hence  $D_f$  will also be low. On the contrary, when the denoising method attenuates noise and retains signal,  $N_f$  and  $S_f$  will be high and hence  $D_f$  will also be high.

Based on this criterion, we chose  $f_t$  as that corresponding to maximum  $D_f$  for frequencies from 1 to 10 Hz. A range from 1 to 10 Hz was chosen because the ERP is known to be a low-frequency waveform [13]:

$$f_t = \max(D_{f=1, \dots, N}). \quad (13)$$

For the simulated dataset,  $f_t$  according to (13) was 1 Hz as can be seen from Table 1. Figure 3 shows a close correspondence between EMD-denoised average and raw average, suggesting that most of the variance due to the signal has been retained in spite of the denoising

**3.1.3. Analysis.** To evaluate the performance of EMD-denoising on simulated data, with  $f_t = 1$  Hz, a number of tests were conducted. The SNR of the dataset with raw trials was compared to the SNR of the denoised dataset. To further investigate denoising performance, we calculated measures reflecting the two aspects of denoising—noise reduction and signal retention.

(a) *Noise Reduction Percentage.* The average variance due to noise on each raw single trial,  $NV_R$ , can be calculated by

TABLE 1: Denoising performance as function of threshold frequency.

Threshold frequency (Hz)	Denoising performance (units)
1	0.133
2	0.129
3	0.119
4	0.115
5	0.110
6	0.107
7	0.104
8	0.101
9	0.099
10	0.097

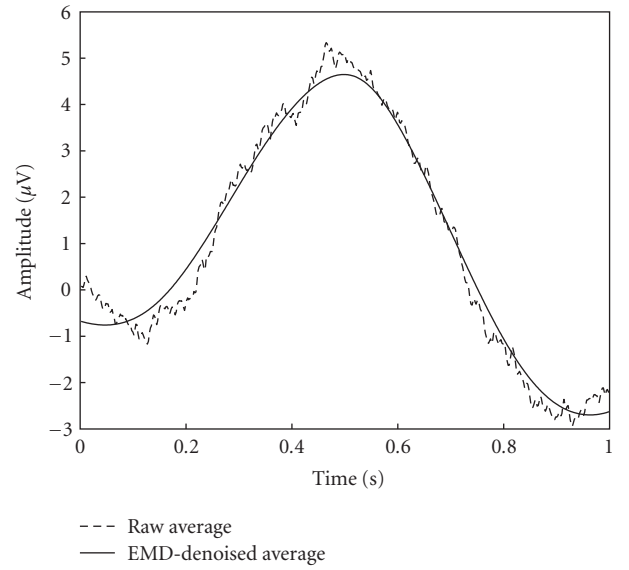


FIGURE 3: EMD-denoised average superimposed on raw average. Threshold frequency was 1 Hz, calculated according to the criterion in (12).

subtracting the ERP from each raw single trial and calculating the average residual variance. The average variance due to noise on each EMD-denoised single trial,  $NV_D$ , can be calculated by subtracting EMD-denoised average from each EMD-denoised single trial and calculating the average variance of the resultant residual. The noise reduction percentage was calculated as

$$\text{NR}\% = 100 \left( 1 - \frac{NV_D}{NV_R} \right). \quad (14)$$

(b) *Signal Retention Percentage.* The signal retention percentage was calculated as the ratio between variance of denoised average,  $X_D$ , to variance of raw average,  $X_R$ , multiplied by 100

$$\text{SR}\% = 100 \left( \frac{\text{Var}(X_D)}{\text{Var}(X_R)} \right). \quad (15)$$

Since the ERP waveform was known for the simulated data, Root Mean Square Error (RMSE), Amplitude deviation, and Latency deviation were used to assess the method's ability to estimate shape, amplitude, and latency of the single-trial ERP, respectively. Since the simulated data had similar properties to experimental ERP data, it was considered that these results would reflect the method's ability to estimate shape, amplitude, and latency of a single trial ERP on experimental data as well.

(c) *Root Mean Square Error (RMSE)*. Since in the simulated data, we have knowledge about the time course of the ERP, we used RMSE to evaluate the extent to which both the EMD-denoised trials and raw single trials accurately represented the ERP. If  $x$  is a single trial from either EMD or raw trials and  $x_{\text{ref}}$  is the reference ERP,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^M (x_i - x_{\text{ref}})^2}{M}}, \quad (16)$$

where  $M$  is the number of samples for each single trial. Distributions corresponding to EMD and raw trials were then compared using either paired-samples  $t$ -test or Wilcoxon signed-rank test depending on whether both distributions were normal or not. Normality was tested for using Lilliefors test. Lower RMSE indicates better estimation of ERP time course.

(d) *Amplitude Deviation*. The extent to which peak amplitude is estimated correctly was measured for each EMD-denoised single trial and raw single trial using the metric of amplitude deviation

$$\sigma_{\text{amp}} = |x_{\text{amp}} - \hat{x}_{\text{amp}}|, \quad (17)$$

where  $x_{\text{amp}}$  is actual peak amplitude while  $\hat{x}_{\text{amp}}$  is the observed peak amplitude. Since this metric was calculated on artificial data,  $x_{\text{amp}}$  was known to be  $5 \mu\text{V}$ .  $\hat{x}_{\text{amp}}$  was calculated using peak-picking, choosing the maximum positive voltage value. Distributions corresponding to EMD and raw trials were then compared using either paired-samples  $t$ -test or Wilcoxon signed-rank test depending on whether both distributions were normal or not. Normality was tested for using Lilliefors test. A smaller amplitude deviation indicates better amplitude estimation.

(e) *Latency Deviation*. The extent to which peak latency is estimated correctly was measured for each EMD-denoised single trial and raw single trial using the metric of latency deviation

$$\sigma_{\text{lat}} = |x_{\text{lat}} - \hat{x}_{\text{lat}}|, \quad (18)$$

where  $x_{\text{lat}}$  is the actual peak latency while  $\hat{x}_{\text{lat}}$  is the observed peak latency.  $x_{\text{lat}}$  was known to be 486 ms while  $\hat{x}_{\text{lat}}$  was calculated as the latency corresponding to the observed peak amplitude. Distributions corresponding to EMD and raw trials were then compared using either paired-samples  $t$ -test or Wilcoxon signed-rank test depending on whether both

distributions were normal or not. Normality was tested for using Lilliefors test. A smaller latency deviation indicates better latency estimation.

3.2. *Experimental Data*. In addition to comparing EMD-denoised trials and raw trials on simulated data, we also evaluated EMD on experimental data.

3.2.1. *Data*. The experimental data was from an ERP study on language processing in the Basque language. Number of participants was 25, all native Basque speakers. Condition A contains grammatically correct sentences and is the control condition. Condition B contains sentences in which the licenser (e.g., negation) of a Negative Polarity Item (NPI) is inaccessible. For each condition, 50 sentences were presented to each subject word by word. At the end of each sentence, participants were asked to indicate whether it was grammatically correct. Participants response to the NPI of each sentence was analysed.

ERP data was recorded on a 64-electrode array. The sampling frequency was 250 Hz. All trials were baseline corrected by subtracting value at each sample by mean of the activity 200 ms prestimulus.

### 3.2.2. EMD Denoising

(a) *Electrode Selection*. The analysis was to be performed on single-channel data, hence it was necessary to choose an electrode which gives good separation between condition A and condition B. To do this, mean effect size between baseline-corrected raw trials from conditions A and B was calculated for each electrode. The mean was calculated across effect sizes for each sample from stimulus onset up to 1 s poststimulus.

Effect size,  $d$ , is calculated according to the formula

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}, \quad (19)$$

where  $x_1$  is average of the first sample of a given variable and  $x_2$  is average of the second sample of the same variable. The pooled standard deviation  $s$  is given by

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2)}}, \quad (20)$$

where  $n_1$  is the number of items in the first sample of standard deviation  $s_1$  and  $n_2$  is the number of items in the second sample of standard deviation,  $s_2$ .

Based on this criterion, the C2P electrode was identified as providing the best separation between condition A and condition B. Subsequent analysis was thus confined to C2P electrode.

(b) *Selecting Threshold Frequency ( $f_t$ )*. It is also necessary to choose a threshold frequency ( $f_t$ ) as parameter for use by EMD. EMD will include all IMFs with dominant frequencies below this threshold frequency, for single-trial reconstruction. The approach to selecting threshold frequency was



similar to that applied on the simulated data as specified in (13). However, since there are two conditions in this case, denoising performance was calculated for each condition at each threshold frequency between 1 and 10 Hz. The selected threshold frequency was the one which had the highest sum of values for denoising performance,  $D_f$ , which was estimated for condition A and condition B separately. The other difference from the selection of  $f_t$  in simulated data was that the 25 single-subject averages were used in this case rather than using single trials. This was because single-subject averages are of higher SNR, thereby potentially giving more accurate estimation of optimal threshold frequency. The selected threshold frequency was 3 Hz for this dataset. As seen in Figure 6, there is a close correspondence between EMD-denoised averages and averages of trials lowpass filtered at 40 Hz.

### 3.2.3. Analysis

(a) *Effect Size.* For experimental data, the actual time course of the ERP is unknown, and so, it is not possible to assess the ability of EMD to estimate time course, amplitude, and latency of the ERP. However, one can compare the ability of EMD to separate the two conditions against that of the conventional preprocessing method of lowpass filtering at 40 Hz. Since effect size is a measure of separation of responses from two conditions, it is employed as a performance measure to compare EMD against this baseline condition of lowpass filtering.

From inspection of Figure 6, it seems that event-related differences extend from time of stimulus onset up to 1 s poststimulus. Thus, for each of the 25 subjects, mean effect size of each sample from stimulus onset to 1 s poststimulus was calculated for both EMD-denoised trials and trials lowpass filtered at 40 Hz. Both EMD-denoised trials and lowpass filtered trials had been baseline corrected with activity 200 ms prestimulus. Distributions corresponding to EMD and lowpass filter were then compared using either paired-samples  $t$ -test or Wilcoxon signed-rank test depending on whether both distributions were normal or not. Normality was tested for using Lilliefors test.

(b) *Comparison of Correct and Incorrect Trials.* On both condition A and condition B, participants correctly responded to questions most of the time while responding incorrectly on rare occasions. Due to the low number of incorrect responses, any differences between correct and incorrect responses will likely be hard to detect. Hence, it presents a scenario suitable for evaluating effectiveness of denoising methods.

For condition A, responses from all 25 participants were included. In all, there were 1010 correct responses (84.7%) and 183 incorrect responses (15.3%). For condition B, only 17 participants were included (subjects 9 to 25) since correct/incorrect information was available for this condition only for these subjects. The number of correct responses was 764 (95.7%), and number of incorrect responses was 34 (4.3%).

The correct and incorrect responses for both conditions were denoised using two methods—EMD and lowpass filtering at 40 Hz. lowpass filtering at 40 Hz is a conventional preprocessing procedure for ERP analysis. Then, correct and incorrect responses from different methods were compared using independent samples  $t$ -test, for each condition separately. This was done on trials both from condition A and condition B that had been baseline corrected with activity 200 ms prestimulus.

## 4. Results

Figure 4 shows an ERP image of the ensemble of raw single-trials while Figure 5 shows an ERP image of an ensemble of EMD-denoised single trials. An ERP image, introduced in [31], is a convenient way of visualising an ensemble of single trials, with time on the  $x$ -axis, trial number on the  $y$ -axis, and normalised amplitude represented as colour. From the smoother ERP image of Figure 5, it is seen that EMD-denoising reveals the ERP much more clearly.

This suggests that EMD-denoised single trials could reveal trial-trial variations in amplitude or latency of the single-trial ERPs as a function of some behavioural or experimental variable. For example, in [6], Jung et al. used an ERP image to reveal that the peak latency of the P3 ERP component from a visual selective attention experiment was a function of reaction time.

4.1. *Simulated Data.* SNR of the denoised dataset was 4.3 compared to 1 for that of the raw dataset. Table 2 gives details about performance of the denoising method on two important criteria. On artificial data, EMD is able to reduce variance due to background EEG by as much as 77% and retains almost all variance due to the ERP (99.5%).

At least one of the two distributions (corresponding to EMD and raw trials) was found to be significantly nonnormal according to Lilliefors test for RMSE, amplitude deviation, and latency deviation comparisons. Hence, Wilcoxon signed-rank test was used for all comparisons.

EMD performs significantly better than the baseline condition of raw single trials for RMSE, amplitude deviation, and latency deviation. As shown in Figure 7, EMD (mean = 1.33) has significantly lower RMSE than raw single trials (mean = 2.53),  $Z = -8.7$ ,  $P < .001$ .

EMD also has lower amplitude deviation (mean = 1.19) than raw single trials (mean = 4.57)  $Z = -8.5$ ,  $P < .001$ . Further, it has lower latency deviation (mean = 0.04) than raw single trials (mean = 0.07),  $Z = -5$ ,  $P < .001$ , as shown in Figure 9. Thus, EMD gives a significantly better estimate of the shape, peak amplitude, and peak latency of the ERP compared to raw single trials. In terms of decibels, the performance improvement due to EMD-denoising is  $-5.6$  dB for RMSE and  $-11.7$  dB for amplitude deviation, with negative values due to denoised values being lower than raw values. Performance improvement of latency deviation cannot be expressed in decibels as it is measured in ms.

4.2. *Experimental Data.* On experimental data, EMD-denoising was compared with lowpass filtering at 40 Hz

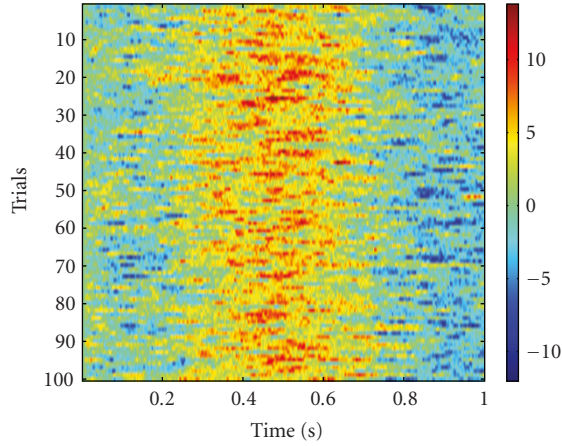


FIGURE 4: ERP image of ensemble of raw single trials.

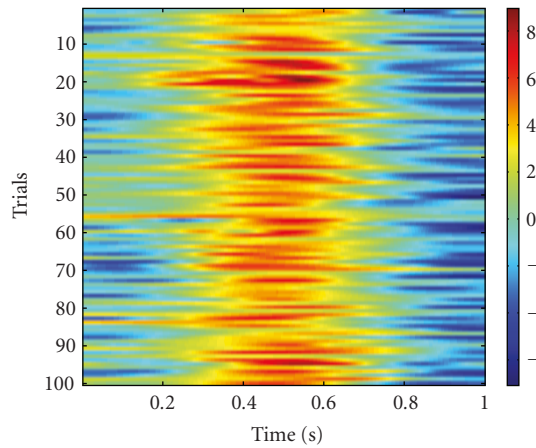


FIGURE 5: ERP image of ensemble of EMD-denoised single trials.

TABLE 2: Denoising performance of EMD.

Noise reduction (%)	Signal retention (%)
77	99.5

for its ability to differentiate between two conditions, as measured by effect size. According to Lilliefors test, both distributions (corresponding to EMD and lowpass filtering) are normal distributions. Hence, paired-samples  $t$ -test was used for the comparison. As illustrated in Figure 10, EMD-denoised trials (mean = 0.33) were found to have a higher effect size than trials lowpass filtered at 40 Hz (mean = 0.26),  $t(24) = 5.15$ ,  $P < .001$  suggesting that it is able to better discriminate between conditions. Performance improvement in effect size due to EMD-denoising cannot be expressed in decibels as it is a dimensionless quantity.

When comparing correct and incorrect trials on condition A, EMD reveals a significant difference in a time window from 436 ms to 528 ms poststimulus, as shown in Figure 11. On condition B, EMD-denoised trials reveal significant differences between correct and incorrect trials

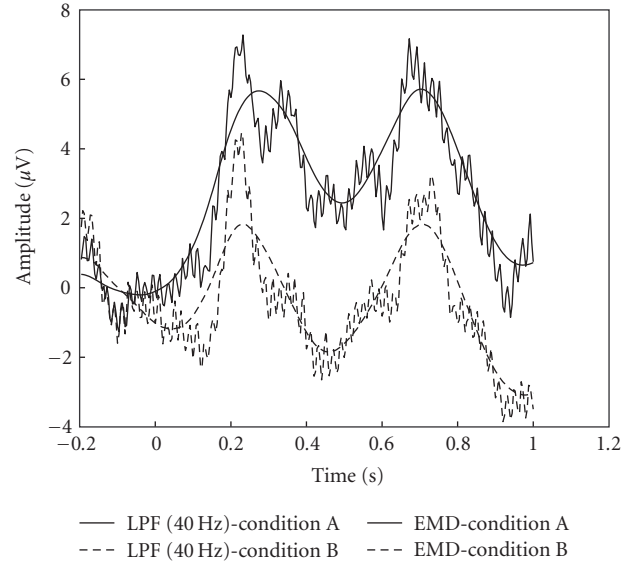


FIGURE 6: EMD-denoised averages for conditions A and B are superimposed onto corresponding averages of trials lowpass filtered at 40 Hz. Threshold frequency is 3 Hz, calculated according to the criteria in (13).

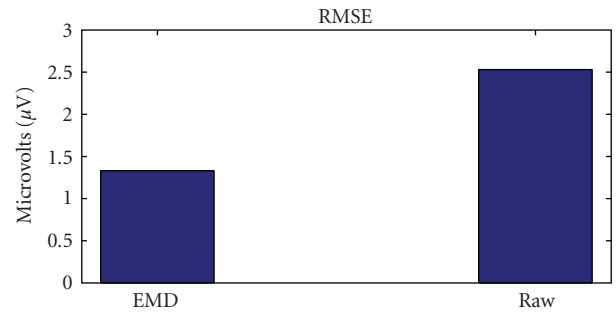


FIGURE 7: Bar graph comparison of RMSE for EMD-based and raw single trials. Lower RMSE indicates better performance.

from 348 ms to 616 ms and again from 884 ms to 1000 ms, as shown in Figure 12.

Figure 13 shows the comparison for trials lowpass filtered at 40 Hz for condition A. The windows of significance occur sporadically and are of much shorter length, none longer than 20 ms (488 ms to 508 ms). For the equivalent EMD comparison, the longest window of significance is 92 ms (from 436 ms to 528 ms). For trials lowpass filtered at 40 Hz on trials from condition B, clear differences are revealed between correct and incorrect trials as shown in Figure 14. However, the windows of significance also occur sporadically and are of shorter duration than for the equivalent EMD comparison. The window of longest duration for trials lowpass filtered at 40 Hz was 112 ms (from 436 ms to 548 ms) while it was 268 ms (from 348 ms to 616 ms) for the equivalent EMD comparison.

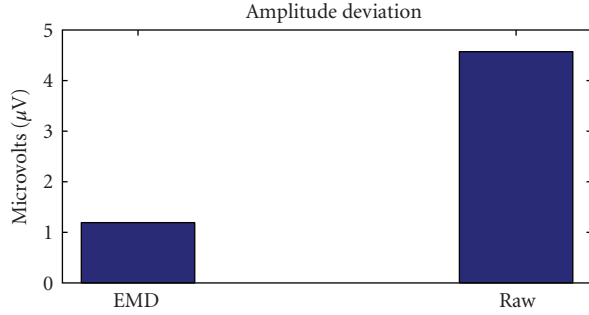


FIGURE 8: Bar graph comparison of amplitude deviation for EMD-based and raw single trials. Lower amplitude deviation indicates better performance.

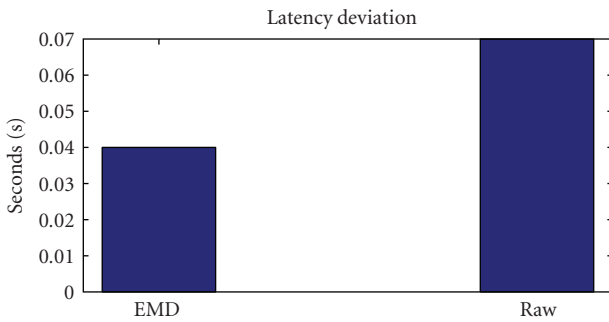


FIGURE 9: Bar graph comparison of latency deviation for EMD-based and raw single trials. Lower latency deviation indicates better performance.

## 5. Discussion

EMD is a promising data driven method for denoising ERP data. It has an advantage over other denoising methods because its data driven nature allows it to decompose a time series according to intrinsic time scales. Specifically, the decomposition is done according to the local characteristic temporal scale, and hence, EMD can estimate signals which are nonstationary in nature.

**5.1. Simulated Data.** On artificial data, EMD showed the ability to reduce noise substantially while retaining the signal. It also showed the ability to estimate the shape, amplitude, and latency of the ERP with low error. Its ability to estimate the amplitude and latency of the ERP with low error on a single-trial basis lends itself to being used to track trial-trial variations in the amplitude and/or latency of the ERP which could then be related to some aspect of cognitive processing. Some examples of these applications are given in [32]. The ability to estimate latency more accurately than from raw trials can also be used to improve performance of latency-corrected averaging proposed by Woody in [33].

**5.2. Experimental Data.** EMD was also shown in our study to increase the effect size between two conditions, showing that it can be used to increase the discriminability between two different conditions. This ability translates to requiring

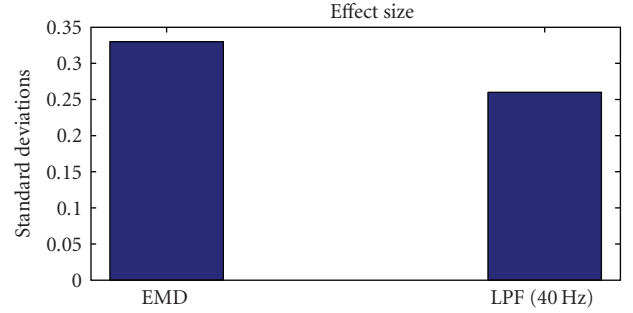


FIGURE 10: Bar graph comparison of effect size for EMD-based and trials lowpass filtered at 40 Hz. Higher Effect size indicates better performance.

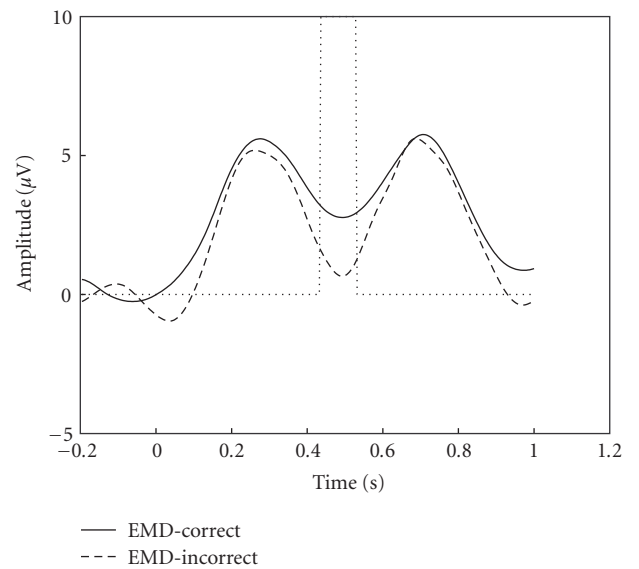


FIGURE 11: Comparison between averages of correct and incorrect EMD-denoised trials, from condition A. Dotted vertical line indicates time points at which difference between two conditions was significant ( $P < .05$ ).

fewer stimulus presentations to obtain a significant difference between two conditions. This in turn could lead to shorter experimental times and less participation fatigue. The level of denoising offered by EMD also makes it possible to conduct experiments and analyse data from single subjects rather than averaging across large groups of subjects. This opens up the possibility of investigating brain dysfunction in brain-impaired populations on a single-subject basis. The noise reduction offered by EMD should also make it possible to clearly detect effects with only a few trials that were hitherto detected only weakly or not detected at all using conventional processing procedures. One example of this is the differences in event-related activity revealed by EMD on condition A, between correct and incorrect trials, when subjects were asked to assess grammatical correctness of a sentence. Another potential application is a comparison between responses categorised according to reaction time as “fast” and “slow.”



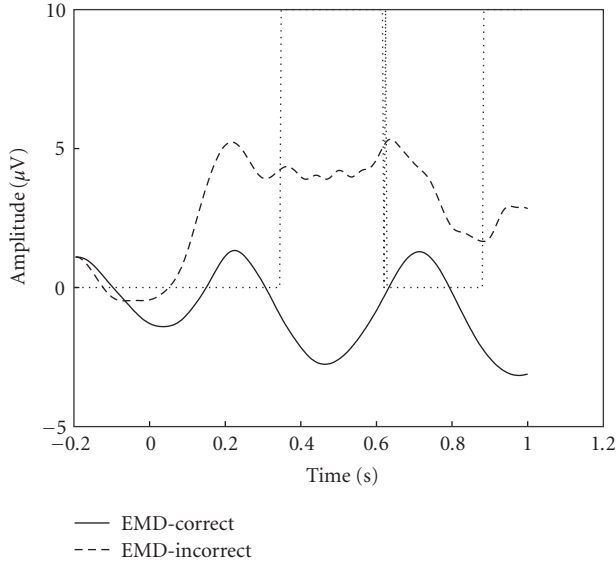


FIGURE 12: Comparison between averages of correct and incorrect EMD-denoised trials, from condition B. Dotted vertical line indicates time points at which difference between two conditions was significant ( $P < .05$ ).

**5.3. Physical Meaningfulness of EMD-Based Decomposition.** The ability of EMD to decompose the data in such a way that small groups of IMFs are event related might or might not be because individual IMFs are physically meaningful. It might be that since EMD tends to decompose a signal into IMFs of progressively lower frequencies and the ERP contains mostly low frequency energy, forming a partial reconstruction from just low-frequency IMFs would naturally perform a denoising function. Hence, the good performance of EMD need not be due to physical meaningfulness of IMFs. There is reason to expect, however, that the EMD decomposition does in fact produce physically meaningful IMFs. EMD assumes that an input time series is a superposition of intrinsic oscillatory modes. This incidentally corresponds to our understanding of EEG/ERP data—that oscillations of different frequency ranges (alpha, beta, gamma, delta, and theta) are functionally distinct and superimposed on each other. In fact, Wang et al. showed in [34] that when EMD was applied to EEG data and the IMFs were clustered, distinct clusters were formed whose frequency spectra corresponded to the naturally occurring frequency bands mentioned above. However, to fully understand whether the EMD decomposition of ERP data is physically meaningful or not, further investigations on simulated should be conducted.

**5.4. Limitations of EMD-Based Denoising.** The proposed denoising methods does have its limitations. One of these is that it is not parameter-free, but rather requires the setting of a parameter, the threshold frequency ( $f_t$ ). The decomposition also does not lend itself to being guided according to a priori knowledge, within a regularization framework, as is the case for the ICA-based decomposition proposed in [35].

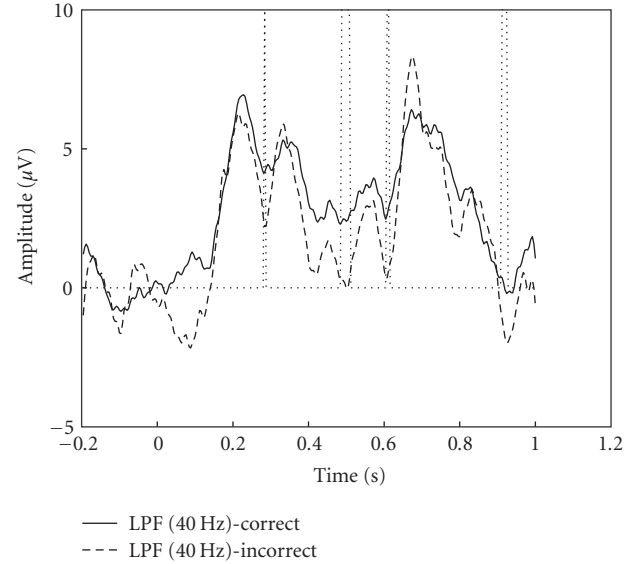


FIGURE 13: Comparison between averages of correct and incorrect trials lowpass filtered at 40 Hz, from condition A. Dotted vertical lines indicate time points at which difference between two conditions was significant ( $P < .05$ ).

Further, the EMD technique by itself has some drawbacks. Its theoretical foundations are not yet well understood, hence making it difficult to interpret the results of decomposition within any formal theoretical framework. It also suffers from border effects which can propagate towards the center especially when trial lengths are short, as mentioned in [24]. The efficacy of the decomposition also depends to a large extent on the suitability of the interpolation scheme which forms the upper and lower envelopes. Cubic spline interpolation is used most often, but other types have also been tried, sometimes with poor results. The stopping criterion for the “sifting process” has also been the subject of much debate, and a number of approaches have been proposed. Also, because of its high dependence on identifying extrema, a certain amount of oversampling is required. Finally, it has been shown that when two oscillatory modes are superimposed with certain combinations of frequencies and relative amplitudes, EMD is not able to separate them [27]. In spite of these limitations though, EMD does decompose the data into physically meaningful components—the primary reason for growing interest in this technique.

**5.5. Comparing with Previous EMD-Based Studies.** It is useful to compare our study with those by [24, 25]. In that case, a multichannel version of EMD has been used but more importantly, they have applied EMD to a special class of evoked potentials, called SSVEPs (steady-state visual evoked potentials). SSVEPs are different from conventional ERPs in a number of ways. For example, SSVEPs are periodic and stationary, and the method for choosing event-related IMFs exploits prior knowledge about this periodic, stationary character of SSVEPs. By contrast, when applying

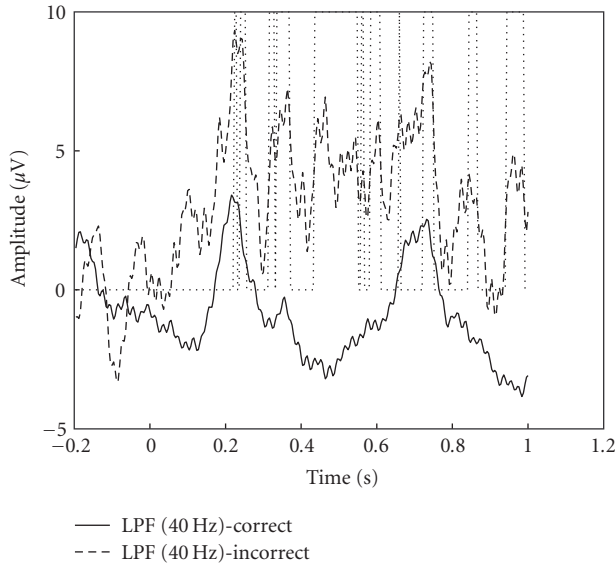


FIGURE 14: Comparison between averages of correct and incorrect trials lowpass filtered at 40 Hz, from condition B. Dotted vertical line indicates time points at which difference between two conditions was significant ( $P < .05$ ).

EMD to single-channel ERP analysis, we do not make any assumptions about periodicity or stationarity of the signal. In fact, EMD-based denoising provides a potential benefit compared to other methods because it allows for non-stationarity of the signal. Thus, although our method and the method described in [24, 25] both use EMD, they are used in different ways to solve different types of filtering problems.

**5.6. Future Work.** As a next step, it is crucial to compare the denoising performance of EMD with other methods, especially nonstationary methods like that based on the Wavelet Transform (WT), recently proposed by Quian Quiroga and Garcia [11]. It is also important to consider alternative ways of selecting event-related IMFs, for example, by using their statistical properties as proposed in [19] or by using measures such as Cumulative Mean Square Error (CMSE) as proposed in [20, 21]. Finally, investigations into the functional significance of IMFs might be useful, for example, by assessing the ability of EMD to reveal event-related effects on higher-frequency IMFs not normally included in the EMD-denoised single trial.

## 6. Conclusion

EMD is a promising data driven, nonstationary method for estimating ERPs. For single-trial analysis, it could be used to study trial-trial variations in amplitude and/or latency of the ERP and relate it to some aspect of cognitive processing. For few-trial analysis, and it can be used to detect effects by comparing averages of only a few EMD-denoised trials. Often, these effects would not be revealed by averaging a few trials that have been processed by conventional means. For the full potential of EMD to be exploited, however, it is

important to understand the functional significance of the IMFs, thereby potentially facilitating our understanding of event-related dynamics in the context of ongoing EEG.

## Acknowledgments

The authors would like to thank the reviewers whose helpful and insightful comments greatly improved the quality of the paper. The authors also thank the University of Reading, UK, for the Research Endowment Trust Fund (RETF) studentship which supported this study. The authors are grateful to Dr. Leticia Pablos Robles for providing access to data from the Basque Language Study.

## References

- [1] M. D. Rugg and M. G. H. Coles, *Electrophysiology of Mind, Event-Related Brain Potentials and Cognition*, Oxford University Press, Oxford, UK, 1994.
- [2] S. J. Luck, *An Introduction to the Event-Related Potential Technique*, MIT, Cambridge, Mass, USA, 2005.
- [3] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.
- [4] S. Cerutti, G. Baselli, D. Liberati, and G. Avesi, "Single sweep analysis of visual evoked potentials through a model of parametric identification," *Biological Cybernetics*, vol. 56, pp. 111–120, 1987.
- [5] H. J. Heintze and H. Konkel, "ARMA-filtering of evoked potentials," *Methods of Information in Medicine*, vol. 23, pp. 29–36, 1984.
- [6] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Analysis and visualization of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 166–185, 2001.
- [7] R. Vigario and E. Oja, "Independent component analysis of human brain waves," in *Engineering Applications of Bio-Inspired Artificial Neural Networks. Proceedings of the 10th International Work-Conference on Artificial and Natural Neural Networks (IWANN '99)*, vol. 2, pp. 238–247, Alicante, Spain, June 1999.
- [8] R. M. Chapman and J. W. Mccrory, "EP component identification and measurement by principal components-analysis," *Brain and Cognition*, vol. 27, no. 3, pp. 288–310, 1995.
- [9] D. D. B. D. Rubin, G. Baselli, G. F. Inbar, and S. Cerutti, "An adaptive neuro-fuzzy method (ANFIS) for estimating single-trial movement-related potentials," *Biological Cybernetics*, vol. 113, pp. 63–75, 2002.
- [10] D. O. Walter, "A posteriori "Wiener filtering" of average evoked responses," *Electroencephalography and Clinical Neurophysiology*, vol. 27, supplement, pp. 61–70, 1968.
- [11] R. Quian Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clinical Neurophysiology*, vol. 114, no. 2, pp. 376–390, 2003.
- [12] J. P. C. de Weerd and J. I. Kap, "A posteriori time-varying filtering of averaged evoked potentials. II. Mathematical and computational aspects," *Biological Cybernetics*, vol. 41, no. 3, pp. 223–234, 1981.
- [13] E. A. Bartnik, K. J. Blinowska, and P. J. Durka, "Single evoked potential reconstruction by means of wavelet transform," *Biological Cybernetics*, vol. 67, no. 2, pp. 175–181, 1992.

- [14] A. Effern, K. Lehnertz, T. Grunwald, G. Fernández, P. David, and C. E. Elger, "Time adaptive denoising of single trial event-related potentials in the wavelet domain," *Psychophysiology*, vol. 37, no. 6, pp. 859–865, 2000.
- [15] W. Klimesch, G. Pfurtscheller, and H. Schimke, "ERD-attentional and cognitive processes in the upper and lower alpha band," *Electroencephalography and Clinical Neurophysiology*, vol. 87, p. 133, 1993.
- [16] J. I. Salisbury and M. Wimbush, "Using modern time series analysis techniques to predict ENSO events from the SOI time series," *Nonlinear Processes in Geophysics*, vol. 9, pp. 341–345, 2002.
- [17] N. E. Huang, M.-L. Wu, W. Qu, S. R. Long, and S. S.P. Shen, "Applications of Hilbert-Huang transform to non-stationary financial time series analysis," *Applied Stochastic Models in Business and Industry*, vol. 19, no. 3, pp. 245–268, 2003.
- [18] A. Andrade, *Decomposition and analysis of electromyographic signals*, Ph.D. dissertation, School of Systems Engineering, University of Reading, Reading, UK, 2005.
- [19] P. Flandrin, P. Goncalves, and G. Rilling, "Detrending and denoising with empirical mode decomposition," in *Proceedings of the European Signal Processing Conference (EUSIPCO '04)*, pp. 1581–1584, 2004.
- [20] A. O. Boudraa and J. C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2196–2202, 2007.
- [21] A. O. Boudraa and J. C. Cexus, "Denoising via empirical mode decomposition," in *Proceedings of the 2nd IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing (ISCCSP '06)*, pp. 1–4, Marrakech, Morocco, March 2006.
- [22] H. Liang, S. L. Bressler, E. A. Buffalo, R. Desimone, and P. Fries, "Empirical mode decomposition of field potentials from macaque V4 in visual spatial attention," *Biological Cybernetics*, vol. 92, no. 6, pp. 380–392, 2005.
- [23] C. M. Sweeney-Reed and S. J. Nasuto, "A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition," *Journal of Computational Neuroscience*, vol. 23, no. 1, pp. 79–111, 2007.
- [24] T. M. Rutkowski, A. Cichocki, T. Tanaka, A. L. Ralescu, and D. P. Mandic, "Clustering of spectral patterns based on EMD components of EEG channels with applications to neurophysiological signals separation," in *Advances in Neuro-Information Processing*, M. Köppen, N. Kasabov, and G. Coghill, Eds., vol. 5506 of *Lecture Notes in Computer Science*, pp. 453–460, Springer, Berlin, Germany, 2009.
- [25] T. M. Rutkowski, D. P. Mandic, A. Cichocki, and A. W. Przybyszewski, "EMD approach to multichannel EEG data the amplitude and phase components clustering analysis," *Journal of Circuits, Systems and Computers*, vol. 19, no. 1, pp. 215–229, 2010.
- [26] L. Zou, C. Tao, X. Zhang, and R. Zhou, "Estimation of event related potentials using wavelet denoising based method," in *Proceedings of the 7th International Symposium on Neural Networks*, vol. 6064 of *Lecture Notes in Computer Science*, pp. 400–407, Shanghai, China, 2010.
- [27] G. Rilling, P. Flandrin, and P. Gonçalves, "On empirical mode decomposition and its algorithms," in *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03)*, Grado, Italy, June 2003.
- [28] S. Hanslmayr, W. Klimesch, P. Sauseng et al., "Alpha phase reset contributes to the generation of ERPs," *Cerebral Cortex*, vol. 17, no. 1, pp. 1–8, 2007.
- [29] S. Y. Tseng, R. C. Chen, F. C. Chong, and T. S. Kuo, "Evaluation of parametric methods in EEG signal analysis," *Medical Engineering and Physics*, vol. 17, no. 1, pp. 71–78, 1995.
- [30] "Data contributed by Dr. Vince Calhoun," <http://mlsp2005.conwiz.dk/index.php?id=30.html>.
- [31] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends in Cognitive Sciences*, vol. 8, no. 5, pp. 204–210, 2004.
- [32] R. Q. Quiroga, M. Atienza, J. Cantero, and M. Jongsma, "What can we learn from single-trial event-related potentials?" *Chaos and Complexity Letters*, vol. 2, no. 2-3, pp. 345–363, 2006.
- [33] C. D. Woody, "Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals," *Medical & Biological Engineering*, vol. 5, no. 6, pp. 539–554, 1967.
- [34] Z. Wang, A. Maier, N. K. Logothetis, and H. Liang, "Single-trial classification of bistable perception by integrating empirical mode decomposition, clustering, and support vector machine," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 592742, 8 pages, 2008.
- [35] S. Lemm, G. Curio, Y. Hlushchuk, and K. R. Müller, "Enhancing the signal-to-noise ratio of ICA-based extracted ERPs," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 601–607, 2006.