Wavelet-based Semblance for P300 Single-trial Detection

Carolina Saavedra, Laurent Bougrain

Wavelet-based Semblance for P300 Single-trial Detection

Carolina Saavedra[1,2] and Laurent Bougrain[1,2]
[1]*Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France*
[2]*Inria, Villers-lès-Nancy, F-54600, France*
{*carolina.saavedra,laurent.bougrain*}*@loria.fr*

Keywords: Event-Related Potential, Denoising, Wavelets, Signals correlation, Single-trial detection, Brain-Computer Interfaces.

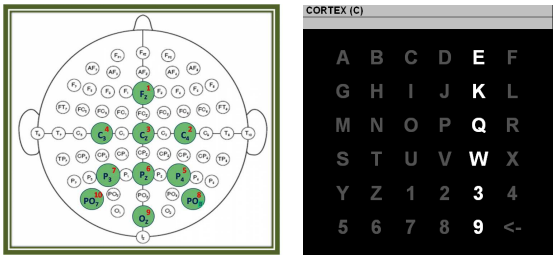
Abstract: Electroencephalographic signals are usually contaminated by noise and artifacts making difficult to detect Event-Related Potential (ERP), specially in single trials. Wavelet denoising has been successfully applied to ERP detection, but usually works using channels information independently. This paper presents a new adaptive approach to denoise signals taking into account channels correlation in the wavelet domain. Moreover, we combine phase and amplitude information in the wavelet domain to automatically select a temporal window which increases class separability. Results on a classic Brain-Computer Interface application to spell characters using P300 detection show that our algorithm has a better accuracy with respect to the VisuShrink wavelet technique and XDAWN algorithm among 22 healthy subjects, and a better regularity than XDAWN.

1 INTRODUCTION

The analysis of brain activity with appropriate techniques allows to extract properties of underlying neural activity and to better understand high level functions. Wavelet are efficient to process non-stationary signals and can be useful to detect event-related potentials (ERP) as the ones used in brain-computer interface (BCI) systems.

A Brain-Computer Interface enables users to act on either a real or a virtual environment by transcribing brain activity into commands for a computer application or other devices (Wolpaw et al., 2002).

The P300 speller (Farwell and Donchin, 1988) is a well-known BCI system which allows the user to write text. It records the brain activity using an electroencephalographic (EEG) system with several electrodes (channels) placed on the scalp (Fig. 1(a)). It uses an oddball paradigm in which low-probability target items are inter-mixed with high-probability non-target items. The speller matrix is usually composed by 36 alphanumeric characters (Fig. 1(b)). Thus, to spell one character it is necessary to flash in random order the 6 columns and the 6 rows, while the user pays attention to the desired letter. When the letter is highlighted, a P300 is generated by the brain. The P300 component is a positive deflection waveform observed around 300 ms after the onset of the stimulus.



(a) (b)

Figure 1: (a) Location of the 10 recorded EEG channels used to detect ERP components; (b) A 6x6 P300 speller.

Thus, the task of the P300 speller system is to detect Event-Related Potentials components from the noisy EEG background signal. It is known to be difficult to accomplish this based on a single trial, because the magnitude of the EEG background activities is usually one-order larger than the one of the ERP components. Moreover, non-invasive electrodes produce a noisy signal because the skull dampens signals. So, the experimental task (single-trial) is repeated many times and the resulting brain activity is averaged over trials to increase the signal-to-noise ratio (SNR). This averaging is made for two reasons: first, the amplitude of the ERP waveform is too small to successfully isolate it from the ongoing EEG activity, and second, it allows the filtering out of artifacts and noise in the signal.

However, three major problems with averaging

ERPs signals are: i) the latency jitter in trials can smooth out the ERP when averaged, ii) fake ERPs due to "phase artifacts", and iii) the communication bit-rates transfer of the system decreases due to the number of repetitions required.

For reducing the number of repetitions, it is necessary to develop robust techniques based on stable features by investigating the time and frequency domains of brain signals.

2 BACKGROUND

There are two techniques commonly used for ERP feature extraction: "Fisher Spatial Filters" whose main objective is finding a linear combination of features which characterize or separate classes (Hoffmann et al., 2006), and "XDAWN", which enhance P300 evoked potentials by using spatial filters based on the *signal to signal plus noise ratio* (Rivet et al., 2009). Despite the success of these techniques, they do not consider the available spectral information in the signal.

Wavelets are very popular for noise filtering of non-stationary signals. They have being used for single-trial ERPs detection in several studies (Quiroga and Garcia, 2003; Yong et al., 2005). The most suitable wavelet denoising technique for EEG signals is the *SureShrink* (Stein Unbiased Risk Estimator) (Donoho and Johnstone, 1995), because it ensures the closest possible reconstruction of the informative signal, minimizing an estimate of the mean square error.

The problem with these proposed solutions using wavelets, is that they denoise one channel at a time, regardless the available information in others channels, not considering the information provided by the ensemble (such as, phase and amplitude information). For this reason this paper presents a new method to denoise EEG signals, which considers the shared information in the wavelet domain of all channels, based on their phase angles correlations. Also, our algorithm is able to select an appropriate temporal window for each subject, extracting the interval of interest to effectively discriminate between classes.

2.1 Wavelets

The Wavelets Transform (WT) is a windowing technique with variable regions size (Mallat, 2008). The main idea is to represent a signal $x(t)$ in terms of displaced and shifted versions of a mother wavelet $\psi(t)$, function:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \tag{1}$$

where a and b are the scale and translation (time shift) parameters respectively.

The signal coefficients are obtained by the convolution of the original signal $x(t)$ and the different versions of the mother wavelet :

$$W_\psi^x(a,b) = \left\langle x(t) | \psi_{a,b}(t) \right\rangle . \tag{2}$$

The coefficients refer to the similarity between the signal and the wavelet at the current scale and time position. It is possible to distinguish two types of wavelet analyses depending on values used to compute the scale a and the translation b: the Continuous Wavelet Transform (CWT), where a and b are continuous and the Discrete Wavelet Transform (DWT), where the discrete orthogonal decomposition is obtained using a discretized scale $a_j = 2^j$ (dyadic step). The time shift b_j is obtained such as that, on a given scale $j+1$, there are twice less coefficients than on the previous scale.

2.2 Wavelet-based Semblance

Semblance analysis was introduced in Geosciences by (Cooper and Cowan, 2008) to compare two given signals $x(t)$ and $y(t)$, using CWT or DWT, based on phase correlations between the wavelet decompositions W_ψ^x and W_ψ^y. The first step is to compute the cross-wavelet transform (Torrence and Compo, 1998):

$$W_\psi^{x,y} = W_\psi^x W_\psi^{y*} \tag{3}$$

where, $*$ denotes the complex conjugate. The cross-wavelet amplitude (also called cross-wavelet power) is given by $A = |W_\psi^{x,y}|$ and its local phase is defined as $\theta = tan^{-1}(\Im(W_\psi^{x,y})/\Re(W_\psi^{x,y}))$, where $\Re$ and $\Im$ correspond to the real and imaginary parts respectively.

The *semblance measure S* can be used to compare two signals using θ, defined as:

$$S = cos^n(\theta), \tag{4}$$

where n is an odd integer greater than zero. The *semblance* measures the correlations between signals based on the scale (wavelength) and time (or position) in the wavelet domain. Its values range from -1 to 1, where $S = 1$ indicates that signals are correlated , $S = 0$ uncorrelated and $S = -1$ inversely correlated. When the mother wavelet is complex, the real and imaginary parts form a Hilbert transform pair, ensuring orthogonality (Cooper and Cowan, 2008).
As equation 4 does not consider the information of the amplitude, it is possible to combine the phase information of S including A as follows:

$$D = cos^n(\theta)|W_\psi^x W_\psi^{y*}|. \tag{5}$$

This measure can be useful if the signal amplitudes are important to be analyzed.

2.3 Wavelet Mean Resultant Length

Because the *semblance* measure is not useful to compare more than two signals, the concept was extended in (Cooper, 2009) based on circular statistics, by considering that the beginning and the end of the phases coincide ($0^\circ = 360^\circ$). The phases can be treated as vectors, because the angles denotes a direction (orientation). By connecting all vectors it is possible to find the *mean orientation* of phases of all signals involved. If the *mean orientation* is divided by the number of vectors used, the Mean Resultant Length (MRL) is obtained. The MRL can be used as a semblance measure for more than two signals, which depends of the number N of signals been treated, so it is possible to compute it for each time t and scale a:

$$MRL(t,a) = \frac{\sqrt{(\sum_{i=1}^{N} \Re(W_\psi^{i,t,a}))^2 + (\sum_{i=1}^{N} \Im(W_\psi^{i,t,a}))^2}}{\sum_{i=1}^{N} |W_\psi^{i,t,a}|} \tag{6}$$

With more than two signals the inversely correlated concept breaks down, which is verified by the MRL values ranging from 0 for uncorrelated signals to 1 for fully correlated signals.

3 METHODS

In this paper we propose to use the *Semblance Measure* and the MRL techniques to denoise brain signals, taking into account the available information of all recorded channels. This is done by using DWT to ensure the reconstruction of the original signal. The choice of DWT is due to its more efficient computation of the inverse Wavelet Transform than CWT. In addition, our algorithm have the ability of localizing the ERP signal in the wavelet domain by selecting an appropriate temporal window adapted to each subject, eliminating non-informative and redundant features.

3.1 Data Denoising

The fundamental hypothesis of wavelet denoising is that wavelets are correlated with the informative signal and not correlated with the noise, which globally means that large coefficients correspond to signal and small coefficients correspond to noise. Therefore, noise canceling can be performed by thresholding (Antoniadis, 2007): only large coefficients will then be used to reconstruct the informative signal.

Let $x_c(t)$ be the signal recorded by the c^{th} channel (or electrode) $c \in \{1,\ldots,C\}$ at time t, $t \in \{1,\ldots,T\}$. The matrix of recorded EEG signals can be defined as $X \in \Re^{TxC}$. The MRL is computed using the wavelet decomposition of all channels $W_\psi^{x_c}, \forall c$ through Equation 6. The MRL coefficients exhibit an exponential distribution, therefore it is possible to establish a correlation threshold τ_d, based on a logarithmic scale, in order to set to zero all coefficients that are below the given threshold. After this process we can reconstruct the signal using the filtered wavelet coefficients.

The MRL computation is made through the combination of the phase angles of the real and imaginary parts of the wavelet decomposition. DWT uses wavelets families that are orthogonal to each other, so the imaginary part can be obtained with the Hilbert transform of the channel. Algorithm 1 shows the denoising process using MRL.

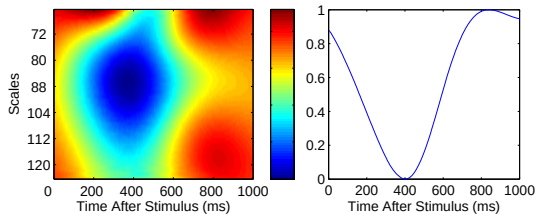
Algorithm 1 MRL Denoising for P300

Input: Given the EEG signal matrix X containing C channels and T temporal samples, and the correlation threshold τ_d (e.g., 0.999)
Output: The denoised signals $\widetilde{X}$

for $c = 1 \rightarrow C$ **do**
  Compute the Hilbert transform H_c of x_c.
  Compute the DWT, $W_\psi^{x_c}$ of signal x_c and $W_\psi^{H_c}$ of H_c using (2).
end for
for $t = 1 \rightarrow T$ **do**
  Compute the $MRL(t)$ using (6).
  if $MRL(t) < \tau_d$ **then**
    set to zero $W_\psi^{x_c}$ at time t, $\forall c$
  end if
end for
for $c = 1 \rightarrow C$ **do**
  Reconstruct the signal for channel c using the new $W_\psi^{x_c}$.
end for

3.2 Window Selection

The common procedure in P300 detection is to study the response during a predefined temporal window after the stimulus onset. Usually this window correspond to 1 second to be sure to include the P300 response and other ERP components in the analysis. However, the P300 responses have a different latency and amplitude for each person, causing data to be included in the temporal window analysis without been a relevant influence in the classifier. We propose to automatically find the temporal window of interest by detecting where the discriminative information lies to remove features which do not carry useful information. The denoised signal can be denoted by $\widetilde{x}_c(t)$, where c correspond to the channel and t to the instant



(a) (b)

Figure 2: (a) Dot product D computed using the complex Morlet wavelet. The colors in the image ranges from blue to red, and passes through the colors cyan, yellow, and orange indicating the similarity between $GA_{\mathcal{T}}$ and $GA_{\mathcal{N}}$ based on the amplitude and phase information ; (b) Average of D normalized between 0 and 1.

when the signal started to be recorded. Each $\widetilde{x}_c(t)$ has a label to indicate to which class belongs. If we denote by $\mathcal{M}$ the set of all signals, $\mathcal{M}$ is composed by signals belonging to the target class $\mathcal{T}$ (containing a P300 wave) and signals $\mathcal{N}$ which are non-targets, $\mathcal{M} = \{\mathcal{T}, \mathcal{N}\}$. The Grand Averages GA for each class are computed as:

$$GA_{\mathcal{T}} = \frac{1}{C|\mathcal{T}|} \sum_{i=1}^{C} \sum_{\widetilde{x} \in \mathcal{T}} \widetilde{x}_i(t) \tag{7}$$

$$GA_{\mathcal{N}} = \frac{1}{C|\mathcal{N}|} \sum_{i=1}^{C} \sum_{\widetilde{x} \in \mathcal{N}} \widetilde{x}_i(t) \tag{8}$$

where the operator $|.|$ denotes the cardinal number. After obtaining the Grand Averages, we compute the CWT $W_{\Psi}^{GA_{\mathcal{T}}}$ and $W_{\Psi}^{GA_{\mathcal{N}}}$ to finally compute the dot product D through equation 5. The result is shown in Figure 2(a). Blue color indicates that the signals have a maximum difference, showing where the P300 is located in the spatial space.

The normalized average avD of D is shown in Figure 2(b). We can see that the P300 has its center around 400 ms. Using a threshold τ_w, $0 \leq \tau_w \leq 1$ the original temporal window of 1s can be reduced to the interval $[t_{lo}, t_{up}]$. Algorithm 2 describes the window selection process.

The combination of algorithms 1 & 2 is called the *Denoise and Window Selection MRL* model (DWS) and take the same temporal window for all channels. A variation on this model is to get a different temporal window for each channel. To do this the only difference in the process is to compute the average by channel instead of the grand averages, and do all the steps for each channel.

Algorithm 2 Window Selection for P300 Algorithm

Input: The denoised signals $\widetilde{X}$ and the correlation threshold τ_w (e.g. 0.9)
Output: The margins for the temporal window t_{lo} and t_{up}

1: Compute the Grand Averages $GA_{\mathcal{T}}$ and $GA_{\mathcal{N}}$ of signals belonging respectfully to the target class using (7) and to the non-target class using (8).
2: Compute the CWTs, $W_{\Psi}^{GA_{\mathcal{T}}}$ and $W_{\Psi}^{GA_{\mathcal{N}}}$ using (2).
3: Compute S, the semblance, using (4) .
4: Compute D using (5).
5: Compute avD, the average of D, and standardize it between 0 and 1.
6: Compute $min(avD)$, the minimum of avD.
7: Compute the lower boundary t_{lo}, the first t to the left of $min(avD)$ which meets $AvD(t) > \tau_w$
8: Compute the upper boundary t_{up}, the first t to the right of $min(avD)$ which meets $AvD(t) > \tau_w$

4 DATABASE

To validate our method, we used a database obtained from first-time users of the P300 speller application implemented within the BCI2000 platform (Schalk et al., 2004). 22 healthy subjects with similar characteristics (sleep duration, drugs, age, etc.) recorded by the Neuroimaging Laboratory of Universidad Autónoma Metropolitana (Mexico) were used. 10 channels, see Figure 1(a), (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz) have been recorded at 256 sps using the g.tec gUSBamp EEG amplifier, a right ear reference and a right mastoid ground. An 8th order Chebyshev bandpass filter, 0.1-60 Hz and a 60 Hz Notch were used. The stimulus is highlighted for 62.5 ms with an inter-stimuli interval of 125 ms.

A complete description of the parameters used for the speller and the data are available in BCI2000 and Matlab formats on the database website: http://akimpech.izt.uam.mx/p300db.

5 EXPERIMENTS

A pre-processing stage were applied to the dataset, prior to the experiments. The data were first filtered by a fourth order forward-backward Butterworth bandpass filter. Cut off frequencies were set to 0.1Hz and 20 Hz (Bougrain et al., 2012). For each channel, the bandpass filtered signals were then normalized having zero mean and a standard deviation equal to one. The temporal window of the post-stimulus response was set to 1 s. For XDAWN, we used the pre-

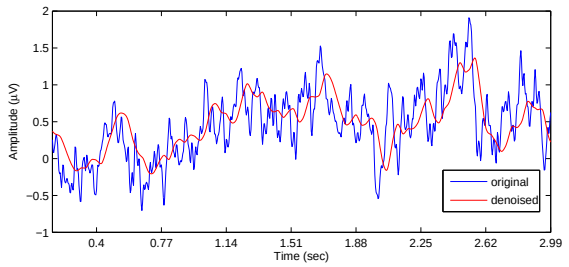


Figure 3: Difference between the original signal and its denoised version. The shown segment correspond to 3 seconds.

processing presented in (Rivet et al., 2009), to obtain the best possible result with this technique.

For the experiments, we used a copy spelling session and a free spelling session of the database respectively for training and testing a linear support vector machine (SVM). The datasets contain 5520 realizations for training and 5895 for testing with a time segment of 1s.

In Figure 3 it is shown the difference between the original signal and the denoised signal after algorithm 1. We note that the signal is best visualized after denoising, making the signal less blurred, what could improve the study and identification of single-trials responses.

In the first experiment we compared ours methods, DWS_1 (with the same temporal window for all channels) and DWS_2 (with different temporal windows per channel) with the wavelet denoising technique called *VisuShrink* and the XDAWN algorithm. In Table 1 the results for 22 subjects are shown. The best result is for the proposed algorithms DWS_1 and DWS_2, proving that the conjoint channel information is useful to deal with the P300 in single-trial problem, obtaining better standard deviations. Moreover as SureShrink and Wavelet-based Semblance are both part of the wavelet theory it is expected that results are similar. However, wavelet-based semblance performs significatively better. The threshold τ_d was set using the formula $1-\tau_d = 10^{-y}$, due to the nature of the distribution of the MRL coefficients. The threshold was tested for values $y = \{1,2,3,4\}$, obtaining the best result for $y = 3$.

Due to the different natures of XDAWN and Wavelet-based Semblance, a more thorough study on each subject was performed to understand the differences in the results. In figure 4 the results by subject are shown, where Wavelet-based Semblance shows better performance in 16 subjects over 22. It is possible to notice that Wavelet-based Semblance and XDAWN have similar results when the subjects have high accuracy rates, which suggest that data is cleaner

Table 1: Results obtained using wavelet family Coiflet level 3, $\tau_d = 0.999$ and $\tau_w = 0.9$. The average and the standard deviation of the letter percentage accuracy over all subjects and the minimum and maximum accuracy obtained among subjects are reported. A paired t-test between DWS1 and XDAWN was significant at a 1% level. A paired t-test between DWS1 and sureShrink was significant at a 5% level.

Method	mean	std	min	max
None	48.23	15.55	18.10	76.19
XDAWN	51.03	15.80	24.44	80.00
Filter [0.1-20]Hz	53.60	14.14	28.25	79.52
SureShrink	54.80	13.90	33.02	78.57
DWS_1	**55.83**	**13.49**	**34.29**	80.95
DWS_2	55.41	13.88	33.97	**81.90**

or the P300 response is stronger making easier the classification. For subject with poor results, Wavelet-based Semblance perform better than XDAWN, therefore we conclude that Wavelet-based Semblance performs better in presence of artifacts. DWS_1 is more stable among subjects XDAWN allows a data reduction by selecting the best combinations of signals. Wavelet-based Semblance achieves a data reduction selecting a shorter temporal window.

Finally, we compare DWS_1 and DWS_2 in Table 2. The lower and upper boundaries obtained for the new

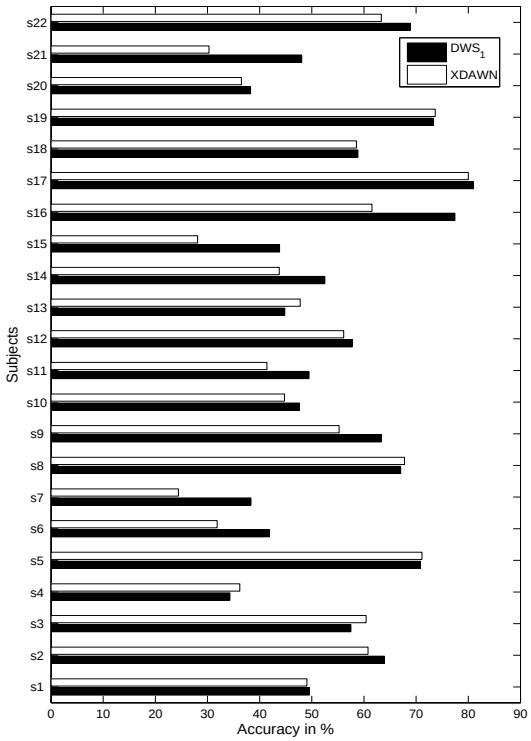


Figure 4: Results for the 22 subjects under study for DWS_1 and XDAWN algorithm.

temporal window are shown, besides the mean, for all subjects of the temporal window size.

For DWS_1 the lower boundary t_{lo} was the highest for subjects 10 and 14, removing approximately 98 ms immediately after the stimulus. This means that this signal segment does not contain yet enough discriminative information to be useful. For the upper boundary t_{up}, the lowest value was obtained by subject 22, removing a little more than half of the original window, because the window was consider until 488 ms to perform the analysis.

Table 2: Results obtained by DWS_1 and DWS_2 on the temporal window selection. t_{lo} and t_{up} are respectively the lower and the higher bounds in ms over all subjects and channels.

		DWS_1	DWS_2
t_{lo}	min	1	1
	mean	20	23
	max	98	305
t_{up}	min	488	277
	mean	848	820
	max	1000	1000

Please recall that, for the algorithm DWS_2, the window selected is different for every channel. The lower boundary t_{lo} was the highest for subject 19 in channel 9 (Oz), suggesting that the information recorded in the first 305 ms does not contain enough discriminative information. The lowest upper boundary t_{up}, obtained for subject 3 at time 277 ms for channel 3 (Cz), which is strange because this channel is usually the one that contains more information. One explanation is that the electrode is not well-fixed to the skull or maybe it was moving at the time of the recording.

6 CONCLUSIONS

Wavelets techniques are becoming an increasingly important exploration tool in BCI, providing temporal and spectral information of the signals under study. In this paper we have introduced a new method to exploit the correlated information among channels based on the wavelet-based semblance, measure that was initially developed to be used in Geosciences. This technique removes the noise and establishes an appropriate temporal window adapted to each subject. Furthermore, the method is quite general and easy to implement, been possible to be used with others brain signals.

We empirically demonstrate using the P300 speller application that our method is useful to remove undesirable component of the signals, improving the letter accuracy for most of the subjects under study. One advantage of this method is the ability to adapt to each subject showing more stability compared with XDAWN. Also, as the number of features is reduced by the window selection algorithm, it is likely that the speed of the classifier may be improved. Despite its advantages, further studies are needed in order to determine the best threshold parameters.

REFERENCES

Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developements and their applications. *Statistics Surveys*, 1:16–55.

Bougrain, L., Saavedra, C., and Ranta, R. (2012). Finally, what is the best filter for p300 detection? *Proceedings of the 3rd TOBI Workshop*, pages 53–54.

Cooper, G. (2009). Wavelet-based semblance filtering. *Computers & Geosciences*, 35(10):1988–1991.

Cooper, G. and Cowan, D. (2008). Wavelet based semblance analysis. *Computers & Geosciences*, 34(2):95–102.

Donoho, D. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224.

Farwell, L. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6):510–523.

Hoffmann, U., Vesin, J.-M., and Ebrahimi, T. (2006). Spatial filters for the classification of event-related potentials. *in Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN)*.

Mallat, S. (2008). *A wavelet tour of signal processing*. Academic Press, 3rd edition.

Quiroga, R. and Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 114(2):376–390.

Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xdawn algorithm to enhance evoked potentials: Application to brain computer interface. *IEEE Trans. Biomed. Engineering*, 56(8):2035–2043.

Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., and Wolpaw, J. (2004). BCI2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043.

Torrence, C. and Compo, G. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791.

Yong, Y., Hurley, N., and Silvestre, G. (2005). Single-trial EEG classification for brain-computer interface using wavelet decomposition. In *European Signal Processing Conference, EUSIPCO 2005*.