Chapter 3
Methodological Framework for EEG Feature Selection Based on Spectral and Temporal Profiles

Vangelis Sakkalis and Michalis Zervakis

Abstract Among the various frameworks in which EEG signal analysis has been traditionally formulated, the most widely studied is employing power spectrum measures as functions of certain brain pathologies or increased cerebral engagement. Such measures may form signal features capable of characterizing and differentiating the underlying neural activity. The objective of this chapter is to validate the use of wavelets in extracting such features in the time–scale domain and evaluate them in a simulated environment assuming two tasks (control and target) that resemble widely used scenarios of assessing and quantifying complex cognitive functions or pathologies. The motivation for this work stems from the ability of time–frequency features to encapsulate significant power alteration of EEG in time, thus characterizing the brain response in terms of both spectral and temporal activation. In the presented algorithmic scenario, brain areas' electrodes of significant activation during the target task are extracted using time-averaged wavelet power spectrum estimation. Then, a refinement step makes use of statistical significance-based criteria for comparing wavelet power spectra between the target task and the control condition. The results indicate the ability of the proposed methodological framework to correctly identify and select the most prominent channels in terms of "activity encapsulation," which are thought to be the most significant ones.

Vangelis Sakkalis
Department of Electronic and Computer Engineering, Technical University of Crete, Chania 73100, Greece, e-mail: sakkalis@ics.forth.gr

Michalis Zervakis
Department of Electronic and Computer Engineering, Technical University of Crete, Chania 73100, Greece, e-mail: michalis@display.tuc.gr

W. Chaovalitwongse et al. (eds.), *Computational Neuroscience*,
Springer Optimization and Its Applications 38, DOI 10.1007/978-0-387-88630-5_3,

3.1 Introduction

Electroencephalographic (EEG) measures have been successfully used in the past as indices of cerebral engagement in cognitive tasks or in the identification of certain brain pathologies. Higher brain functions typically require the integrated, coordinated activity of multiple specialized neural systems that generate EEG signals at various brain regions. Linear [7,18] and nonlinear signal analysis methods have been applied in order to derive information regarding patterns of local and coordinated activity during performance of specific tasks [11] or in various pathologies [2, 13]. The inherent complexity and the dynamic nature of brain function make the evaluation using EEG a rigorous job. Nevertheless, EEG signal analysis provides the advantage of high time resolution and thus it can deduce information related to both local and widespread neuronal activations in short-time periods, as well as their time evolution.

Traditional spectral analysis techniques with Fourier transform (FT) and more specifically the windowed power spectral density function, known as the periodogram [16], form the most commonly used analytical tool for spectral representation and evaluation of activity on different EEG frequency bands [7, 15] – namely *delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ)*. However, this approach considers the EEG signal as a stationary process, which assumption is not satisfied in practice, thus restricting the actual confidence on results. A more promising methodology is based on the time-varying spectral analysis that takes into account the nonstationary dynamics of the neuronal processes [1]. The short-time Fourier (STFT) and the wavelet transforms are the most prevalent analysis frameworks of this class. The first approach uses a sliding time window, whereas the second one forms the projection of the signal onto several oscillatory kernel-based wavelets matching different frequency bands. Currently, such time-varying methods have been widely applied in event-related potential (ERP) data, where distinct waveforms are associated with an event related to some brain function [3]. Under certain assumptions, both time–frequency transforms are in fact mathematically equivalent, since they both use windows that under certain conditions can provide the same results [4]. The reason why these approaches are often regarded as different lies in the way they are used and implemented. Wavelet transform (WT) is typically applied with the relative bandwidth ($\Delta f/f$) held constant, whereas the Fourier approach preserves the absolute bandwidth (Δf) constant. In other words, STFT uses an unchanged window length, which leads to the dilemma of resolution; a narrow window leads to poor frequency resolution, whereas a wide window leads to poor time resolution. Consequently, according to the Heisenberg uncertainty principle one cannot accurately discriminate frequencies in small time intervals. However, the WT can overcome the resolution problem by providing multiresolution analysis. The signal may be analyzed at different frequencies with different resolutions achieving good time resolution but poor frequency resolution at high frequencies and good frequency resolution but poor time resolution at low frequencies. Such a setting is suitable for short duration of higher frequency and longer duration of lower frequency components of the EEG bands. For the purposes of this study the wavelet approach is used.

In this work we attempt to retrieve additional information (as compared to traditional spectral analysis methods) by making use of the time profile of the EEG signal during the target task under study. The motivation for this work stems from the fact that the WT method is able to extract not only the spectral activations but also the time segments at which they occur. It constitutes the cornerstone of our feature extraction scheme and is used for analyzing task-related or control EEG signals by effectively capturing the power spectrum (PS) of each frequency band and channel. In particular, it encodes the activation differences between the mental states of interest. The subsequent feature selection steps apply test statistics on the extracted "time-averaged" PS features. In addition, our approach introduces an extra refinement step that makes further use of the time profile provided by the WT as to derive and encode the temporally activated brain regions and bands. The proposed EEG feature extraction and selection method may also be applied to other similar nonstationary biological signal analysis problems.

3.2 Methods

3.2.1 Methodology Overview

Two different cognitive tasks are assumed for simplicity: the control and the target ones that involve a modulated rather than random activity. The latter task encapsulates the crucial information for extracting both the frequency bands and the location of brain activity, in terms of channel references or groups of channels (related to specific brain areas) as an index of cerebral engagement in certain mental tasks. The testing hypothesis suggests that the target task induces activity on certain brain lobes, reflected on the associated electrodes in a way significantly different compared to a control task. The WT constitutes the cornerstone of feature extraction and is used in analyzing task-related or control EEG signals by effectively capturing the power spectrum (PS) of each band and channel, particularly encoding the activation differences between the tasks. From the technical point of view, statistics is used to extract and select salient features, testing for significance in both the time and scale domains of the signal. The feature selection steps apply test statistics on the extracted "time-averaged" PS features, but in addition our approach introduces an extra refinement step that makes further use of the time profile of the WT, as to derive and encode the temporally activated brain regions and bands. Test statistics form appropriate means for the design of feature selection criteria strictly based on statistical significance; they are simple to implement and often perform better than other heuristic selection methods. To that respect, we base our selection on statistical tests that rely on statistical properties of the feature data under consideration. Hopefully the identified channels and lobes may elucidate any neurophysiological pathways involved in brain function.

A generic overview of the proposed methodology emphasizing the various statistical approaches is illustrated in Fig. 3.1. Different statistical decisions are possible according to the profile of the data under examination. The first choice is based on whether the data are normally distributed, whereas the second is based on the number of different groups under examination – i.e., whether two or more classes (tasks) are being tested. A detailed view of feature selection and refinement blocks matching our data characteristics is presented in Fig. 3.2. The steps involved, as well as their implementation issues, are analyzed in the following sections.

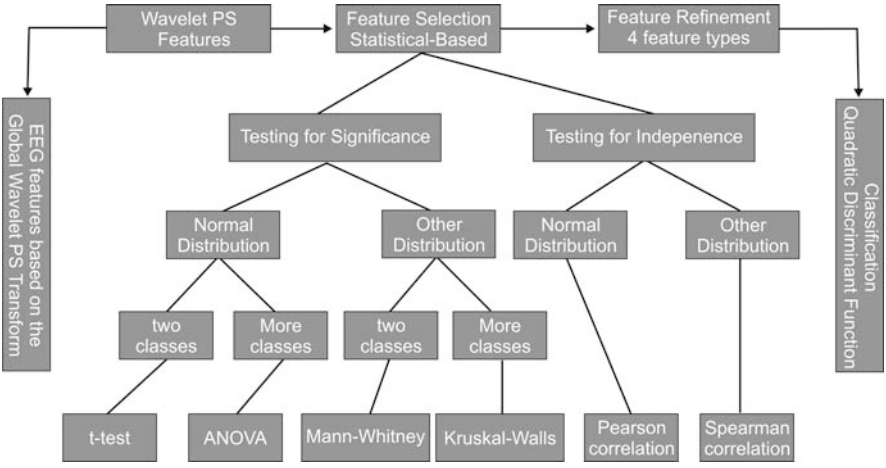

Fig. 3.1: The proposed methodology uses significance-based statistics to reduce the dimensionality of the problem and select the most salient and descriptive feature vectors. Different statistical decisions are possible according to the profile of the data under examination. If one is interested in discriminating two or more classes of normally distributed data, *t*-test or analysis of variance (ANOVA) tests are appropriate candidates, respectively. If the data is nonnormally distributed, Mann–Whitney and Kruskal–Walls tests are the alternatives.

3.2.2 Feature Extraction (Step 1)

Over the past decade the WT has developed into an important tool for analysis of time series that contain nonstationary power at many different frequencies (such as the EEG signal), as well as a powerful feature extraction method [9]. There are several types of wavelet transforms, namely the discrete (DWT) and the continuous (CWT), which involve the use of orthogonal bases or even nonorthogonal wavelet functions, respectively [8]. CWT is preferred in this approach, so that the time and scale parameters can be considered as continuous variables. In the WT, the notion

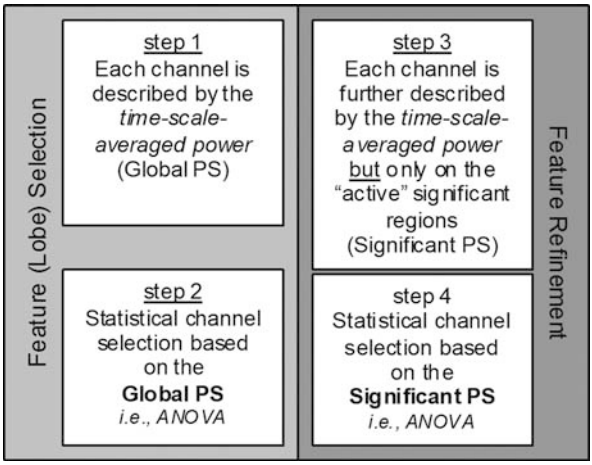

Fig. 3.2: The diagram of the proposed algorithmic transitions, heading toward derivation of significant activity channels and bands.

of scale s is introduced as an alternative to frequency, leading to the so-called time–scale representation domain.

The CWT of a discrete sequence x_n with time spacing δt and N data points ($n = 0, N-1$) is defined as the convolution of x_n with consecutive scaled and translated versions of the wavelet function $\psi_0(\eta)$:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[(n'-n)\delta t/s \right], \tag{3.1}$$

$$\psi_0(\eta) = \pi^{1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{3.2}$$

where η and $\omega_0 = 6$ indicate nondimensional "time" and frequency parameters, respectively and $\psi^*(\cdot)$ denotes the complex conjugate operation. In our application, $\psi_0(\eta)$ describes the most commonly used wavelet type for spectral analyses, i.e., the *normalized complex Morlet wavelet* given in (3.2). The wavelet function ψ_0 is a normalized version of ψ that has unit energy at each scale, so that each scale is directly comparable to each other. The normalization is given as

$$\psi \left[(n'-n)\delta t/s \right] = (\delta t/s)^{1/2} \psi_0 \left[(n'-n)\delta t/s \right]. \tag{3.3}$$

In principle, a complex wavelet function is better suited for capturing oscillatory behavior than a real one, because it captures both the amplitude and the phase of EEG signal. The scale set is given by

$$s_j = s_0 2^{j\delta j}, \; j = 0, \cdots, J, \tag{3.4}$$

where $s_0 = 2\delta t$ is the smallest scale chosen and δj specifies the width of the wavelet function. In our case $\delta j = 0.25$, implying that there is a scale resolution of four sub-octaves per octave [5]. The larger scale is determined by the value of J specified in (3.5), which in our case is $J = 29$:

$$J = \delta j^{-1} \log_2(N\delta t / s_0). \tag{3.5}$$

Finally, the *power spectrum* of the WT is defined by the square of coefficients in (3.1) of the wavelet series as $\|W_n(s)\|^2$. By adopting the above settings a smooth wavelet power diagram is constructed as in Fig. 3.3b for the signal in Fig. 3.3a.

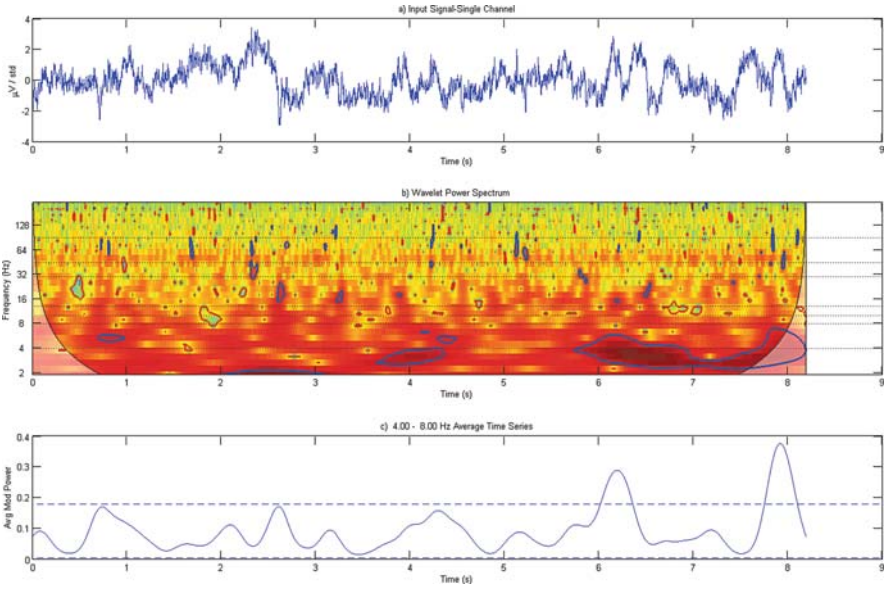

Fig. 3.3: (**a**) A typical normalized EEG signal acquired from a single electrode. (**b**) The wavelet power spectrum presented as a color-coded picture. Mapped scales to frequencies are calibrated on the *y*-axis, with the *horizontal dashed lines* indicating the different frequency bands. The significant regions over the time–scale transform are indicated by *closed contours*. Power increase and decrease is bounded by *blue* and *red* contours, respectively. The outer elliptical region at the edges of this second graph indicates the cone of influence in which errors (edge effects) may be apparent due to the transformation of a finite-length series EEG signal. (**c**) The scalogram of a selected averaged band (Theta 4–8 Hz) reflecting characteristic EEG activity while the participant is performing a complex mathematical calculation [14]. The significance levels are indicated by the *horizontal dashed lines*. PS values greater above the *upper dashed line* indicate significant increase, whereas PS values below the *lower dashed line* indicate significant decrease over the expected control power levels.

As noted before, there exists a concrete relationship between each scale and an equivalent set of Fourier frequencies, often known as *pseudofrequencies* [10]. For the Morlet wavelet used this relationship is $f = \frac{\omega_0+\sqrt{2+\omega_0^2}}{4\pi s}$, which in our case ($\omega_0 = 6$); this gives a value of $f = 1/(1.03s)$. In this study the power spectra is classified in six sequential frequency bands that are coarsely mapped to the scales tabulated in Table 3.1.

Table 3.1: Frequency bands – scale set mapping

Band	Frequency	Scale
Theta (θ)	4–8	21, 22, 23, 24
Alpha1 (α_1)	8–10	20
Alpha2 (α_2)	10–13	18, 19
Beta (β)	13–30	14, 15, 16, 17
Gamma1 (γ_1)	30–45	11, 12, 13
Gamma2 (γ_2)	45–90	7, 8, 9, 10

The first stage of our feature extraction method is based on capturing the *time-averaged power spectrum* $\overline{W_n}^2$ for each electrode and scale, which is computed by averaging the power spectrum $\|\overline{W_n}\|^2$ over time:

$$\overline{W_n}^2(s) = (1/N) \sum_{n=0}^{N-1} \|W_n(s)\|^2. \tag{3.6}$$

Further averaging in scale is performed, in order to map a single feature per frequency band of interest. Thus, the *scale-averaged power spectrum* $\overline{W_n}^2$ is defined as the weighted sum of the wavelet power spectrum $\|W_n(s)\|$ over scales s_{j_1} to s_{j_2} within each frequency band, with scale correspondences defined in Table 3.1. Based on these definitions, the average power over time and frequency band is obtained as

$$\overline{W_{s,n}} = (\delta j/\delta t/C_\delta) \sum_{j=j_1}^{j_2} \left(\|W_n(s_j)\|^2/s_j \right), \tag{3.7}$$

where C_δ is a constant scale-independent factor used for the exact reconstruction of a $\delta(\cdot)$ function from its wavelet transform (for the Morlet wavelet it equals to 0.776) [17]. Once the average PS for each of the studied EEG bands is calculated for each EEG channel and task, we have a high number feature vectors (bands x channels) per task (class), representing each participant (subject), which is actually the time–scale-averaged PS (Global PS – Fig. 3.2 – Step 1) over the band of interest.

3.2.3 Feature Selection (Step 2)

In data mining and classification applications, feature selection and reduction of dimensionality in the feature space play a crucial role in the effective design by regularizing and restricting the solution space. It is of practical concern that a large number of features may actually degrade the performance of the classifier if the number of training samples is limited in comparison to the number of features [12]. This study proposes a statistical method for mining the most significant channels, resembling the way many clinical neurophysiological studies evaluate the brain activation patterns.

Hence, the second step (Fig. 3.2 – Step 2) of our design involves the statistical test selection of features, which depends upon the feature-vector properties and the experimental design. The distribution of features plays the most important role, since it is the one to judge which statistical test is the most appropriate (Fig. 3.1). Normality of the feature set may be tested using the D'Agostino–Pearson test [19]. Once normality is met and supposing that two classes are being discriminated, t-test or analysis of variance (ANOVA) is the ideal test to use in our application. The ANOVA test is superior for complex analyses for two reasons, the first being its ability to combine complex data into one statistical procedure (more than two groups may be compared). The second benefit over a simple t-test is the ANOVA's ability to determine interaction effects. One of the common assumptions underlying ANOVA is that the groups being compared are independent of each other. In the case of a related studies design (the same subjects perform each task), either matched pairs or repeated measures are more appropriate, e.g., a repeated measures ANOVA [19] with common measures factors being the two tasks and the number of channels, testing for significance at the level of 0.05. For those bands where the significance criterion is fulfilled, follow-up post hoc tests for each channel are performed to accentuate the best candidate channels to preserve as features, which resemble the most significant brain areas in terms of activity.

3.2.4 Feature Refinement (Steps 3 and 4)

The aforementioned steps derive a significant channels' subset, based only on task differentiation confidence intervals using Global PS measures. To further refine the features and optimize the whole process, we propose to isolate only those time segments of the EEG signal where notable activity differences occur from the control to the arithmetic task. The aim is to further map the EEG signal into a feature vector that best characterizes the EEG pattern of activity for the target task in terms of significant temporal and spectral content. As we are interested in ongoing EEG activity within various tasks, the temporal activity of EEG events is of interest. Notice that we focus on significant (bursty and/or sequential) activations and not on the evolution of brain operation during the task. Thus, we are mostly focused on the time-localized EEG activity itself, without particular interest to the temporal

relation of these events. We may describe the next step as an attempt to crop up the most significantly different regions from control to target activity out of the bulk initial signal (may be either significant power increase or decrease while performing the requested task compared to the control condition). In fact, this study proposes a way to derive the so-called *significant PS activity* on significantly activated EEG time segments, by testing for significance in the wavelet–time domain the "active" task over the control task (Significant PS – Fig. 3.2 – Step 3). The control task spectra de-fine the *mean time-averaged wavelet power spectrum* over all subjects performing the control task, as

$$\overline{W}(s) = (1/P) \sum_{p=1}^{P} \|W_n^p(s)\|^2, \tag{3.8}$$

where p is the subject index and $W_n^p(s)$ is computed as in (3.1) for each subject. P is the total number of participants. It should be noticed that all EEG signals are normalized to zero mean and identity variance. Further rescaling and comparisons may be performed using each subject's actual signal variance in order to include subject-specific information. Significant power increase on the "active" task is calculated using the 95% confidence level at each scale by multiplying the control task spectrum in Equation (3.8) by the 95th percentile value for a chi-squared distributed variable χ^2 with two degrees of freedom χ_2^2. This is justified because the wavelet power spectrum is derived from the Morlet wavelet in a complex product with the signal, so that both the squares of the real and the imaginary parts of the result are being χ^2 distributed with one degree of freedom each [17, 6]. In a similar manner, significant power decrease is measured using the lower power limit of 5% confidence level at each scale, by multiplying the control task spectrum in Equation (3.8) by the 5th percentile value for the chi-squared distributed variable χ_2^2 . Figure 3.3 depicts one subject's initial normalized EEG signal (Fig. 3.3a) together with its WT (Fig. 3.3b). The significant regions over the time–scale-transformed domain that differentiate the two tasks are indicated by the closed contours; red for significantly increased and blue for decreased activity. Figure 3.3c illustrates another view of the scalogram focusing on a selected averaged band, i.e., (Theta 4–8 Hz). The significance levels in this case are indicated by horizontal dashed lines.

Having derived this significant information, we are now able to form the so-called *significant power spectral* (significant PS) features, which are obtained from the signal energy over those time- and band-localized regions where apparent significant differentiation is indicated (contours in Fig. 3.3b). For the computation of these features, Equation (3.6) is adapted as

$$\overline{W_s t}^2 = (1/m) \sum_{m=m_i}^{m_{i+1}} \|W_m(s)\|^2, i = 1, \cdots, I, \tag{3.9}$$

where m is the total number of time points delimited between the boundaries m_i and m_{i+1} of all significant regions I denoted by each contour in Fig. 3.3b and i is the index of each significant region. Finally, the last step (Fig. 3.2 – Step 4) is actually a repetition of the statistical testing in the second step on the new feature set. ANOVA

or any other better suited statistical method (as described previously) may be used to further sort out and select the best candidate features (significant energy per time, band and electrode), in terms of their task discriminating power.

3.3 Results

The proposed methodology is tested on simulated data, where there exist well-defined spatiotemporal differences in frequency content between the target and the control tasks, as discussed in the following section. In addition, the performance of the proposed approach, as well as its results on actual experimental dataset, is discussed in [14].

3.3.1 Simulation Test

Two different tasks are simulated by two different groups of signals. The first group (control task) consists of 10 simulated spatiotemporal signals, each one comprised of five channels. The idea is to reflect 10 participants virtually registered with a 5-channel-EEG system each. All the channels of the control task are randomly generated quasi-white noise signals, approximately 9-s-long (500 Hz sampling rate – 4,608 samples). The second group (target task) comprises of three channels (channels 1, 3, 4) reflecting white noise and two channels (channels 2, 5) encoding frequency-modulated signals mixed again with quasi-white noise. Channel 2 consists of a time-varying theta EEG signal occurring at a fixed latency, linearly modulated (5–7 Hz) and varying in length randomly between 512 and 1,024 samples among subjects, and a gamma EEG signal, linearly modulated (30–90 Hz) and varying in length randomly between 1,024 and 2,048 samples among subjects, all mixed with quasi-white noise. In a similar manner, channel 5 consists of an alpha band linearly modulated signal (9–12 Hz) varying in length randomly (768–1,536 samples) and a gamma linearly modulated signal (30–90 Hz) varying in length randomly between 512 and 1,024 samples, mixed with quasi-white noise. Quasi-white noise covers the interval between the modulated signals. Such a generated signal (channel 2) together with the wavelet time–frequency representation is depicted in Fig. 3.4. Theta and gamma bands are apparent at different latencies. The tabulated channels in Table 3.2 are the significant ones extracted with the proposed approach from the six (most widely studied) frequency bands (delta, theta, alpha, beta, gamma1, and gamma2). The channels listed in the first column are the selected ones after the first statistical test (Step 2), whereas the channels listed in the second column are the refined ones after the second statistical selection (Step 4). Although the first stage can identify both channels (2 and 5) with the pre-specified frequency content, it is not able to discriminate correctly the activated frequency bands because of leakage effects between bands, as illustrated in Fig. 3.4. In contrast, the second stage focuses

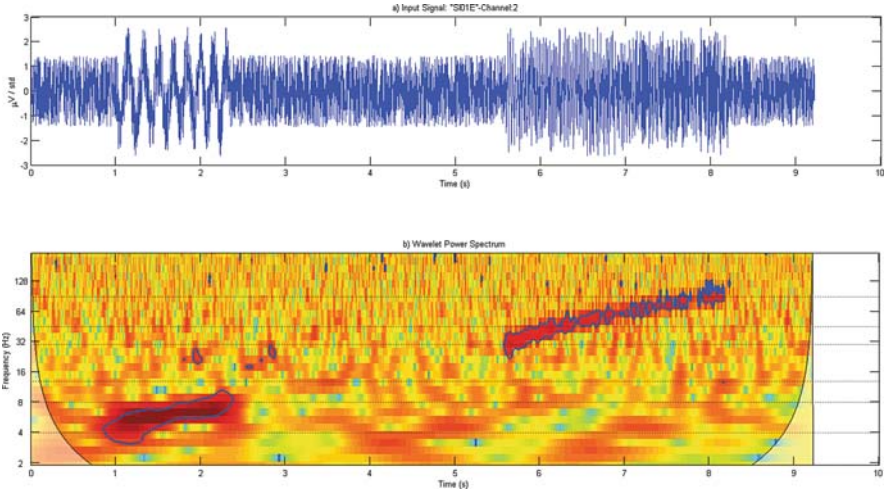

Fig. 3.4: (**a**) The simulated channel 2 consists of a time-varying (among different participants) theta linearly modulated signal (length 2 s) occurring at a fixed latency and a gamma linearly modulated signal (length 3 s) mixed with quasi-white noise. Quasi-white noise also covers the interval between the modulated signals. (**b**) The wavelet PS time–frequency representation picture. The significant regions over the time–frequency transform are indicated by the contours. The significant signal segments (contours) are successfully discriminated from the white noise background.

on the significant regions and is able to detect and correctly account for the energy content of the selected regions.

3.4 Discussion

Using the wavelet transform method on EEG signals, cortical activation evaluation is normally performed by means of comparing a target task (while participant is engaged with a difficult cognitive task or reflects certain pathology) and compares it with a rest condition. This method, in contrast to traditional spectral ones, can estimate changes between EEG signals without being bounded to the stationarity assumption and can provide information for the entire time evolution of the signal.

The simulation test and the results presented justify the suggestion that relevant characteristics are temporally localized in the most significant regions (contours in the WT scalogram), rather than in the entire segment length of the EEGs. The *Global* PS only partially encapsulates the significant information, since there is significant frequency leakage between the bands due to the transient response of the time–frequency filter in different frequencies. Using such features, both channels 2 and 5 in the simulated case induce activity in almost every band. However, the proposed

Table 3.2: Statistical feature – channel selection results

Band	Channel (step 2)	Channel (step 4)	Target
Delta (δ)	2	–	–
Theta (θ)	2, 5	2	2
Alpha (α)	2, 5	5	5
Beta (β)	–	–	–
Gamma1 ($gamma_1$)	2, 5	2, 5	2, 5
Gamma2 ($gamma_2$)	2, 5	2, 5	2, 5

methodology with its second statistical feature selection scheme can efficiently isolate the channels and the correct band activations. Traditional FT spectral analysis methods pose intrinsic limitations on encapsulating the time variation of the signal. Beyond traditional spectral analysis, the WT enables the consideration of time specific significant regions as in Step 3 of the proposed methodology. WT is proved to be a useful measure to detect time-varying spectral power and performs better than traditional time–frequency methods in identifying activity, especially on a shorter temporal scale in high frequencies, which could indicate neuronal synchronous activation in some cortical regions. This is an advantage to previous methodologies, since high-frequency bands are weak and difficult to evaluate using spectral methods.

A qualitative reasoning arising from the application of this methodology to actual data is discussed in [14], where the certain methodology is applied to a complex mathematical reasoning task. Finally, the presented method reveals additional signal characteristics, since it captures not only its average power but also the time-localized activation of the signal.

3.5 Conclusion

The proposed algorithmic approach emphasizes the idea of selecting EEG features based on their statistical significance and further supports the use of time–scale WT domain in order to select significant EEG segments capable of describing the most prominent task-related changes.

Results suggest that the proposed methodology is capable of identifying regions of increased activity during the specified target task. The entire process is automated in the sense that different feature types can be adaptively (according to the data profile) extracted and further refined in a way “transparent” to the user. Such processes may be transferred to a clinical environment if the methods prove to be valuable for the diagnosis of certain pathologies by comparing any routine EEG against a database of pathological ones.

Furthermore, the added value of this approach over other classical Fourier-based methods lies in its ability to further utilize time-domain characteristics of the WT in a way comparable to the evoked potential applications, without making any compromise in the statistical validity of the results.

Acknowledgments This work was supported in part by the EU IST project BIOPATTERN, Contact No: 508803. The Wavelet Transform and various significance testing parts were performed using software implementation based on the wavelet toolbox provided by C. Torrence and G. Compo, available at the URL: `http://paos.colorado.edu/research/wavelets/`.

References

1. Bianchi, A., Mainardi, L., Cerutti, S. Time-frequency analysis of biomedical signals. Trans Inst Measur. and Contr **22**, 321–336 (2000)
2. Breakspear, M., Terry, J., Friston, K., Harris, A., Williams, L., Brown, K., Brennan, J., Gordon, E. A disturbance of nonlinear interdependence in scalp EEG of subjects with first episode schizophrenia. NeuroImage **20**, 466–478 (2003)
3. Bressler, S. The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge MA 412-415, (2002)
4. Bruns, A. Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? J Neurosci Methods **137**(2), 321–332 (2004)
5. Burrus, S., Gopinoth, R., Haitao, G. Introduction to Wavelets and Wavelet Transforms. Prentice Hall Inc., Upper Saddle River, NJ (1998)
6. Chatfield, C. The Analysis of Time Series: An Introduction, 4th edn, p. 241. Chapman and Hall, London (1989)
7. Dumermuth, G., Molinari, L. Spectral analysis of the EEG. Some fundamentals revisited and some open problems. Neuropsychobiology **17**, 85–99 (1987)
8. Farge, M.: Wavelet transforms and their applications to turbulence. Annu Rev Fluid Mech **24**, 395–457 (1992)
9. Kalayci, T., Ozdamar, O. Wavelet preprocessing for automated neural network detection of EEG spikes. IEEE Eng Med Biol Mag **14**, 160–166 (1995)
10. Meyers, S., Kelly, B., O'Brien, J. An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. Mon Wea Rev **121**, 2858–2866 (1993)
11. Micheloyannis, S., Sakkalis, V., Vourkas, M., Stam, C., Simos, P. Neural networks involved in mathematical thinking: Evidence from linear and non-linear analysis of electroencephalographic activity. Neurosci Lett **373**, 212–217 (2005)
12. Raudys, S., Pikelis, V. On dimensionality, sample size, classification error and complexity of classification algorithm in pattern recognition. IEEE Trans Pattern Anal Machine Intell, **PAMI-2**(3), 242–252 (1980)
13. Sakkalis, V., Giurcaneanu, C., Xanthopoulos, P., Zervakis, M., Tsiaras, V., Yang, Y., Micheloyannis, S. Assessment of linear and nonlinear synchronization measures for analyzing EEG in a mild epileptic paradigm. Accepted to be published in, 2008. IEEE Trans Inf Tech (2008). DOI: 10.1109/TITB.2008.923141.
14. Sakkalis, V., Zervakis, M., Micheloyannis, S. Significant EEG features involved in mathematical reasoning: Evidence from wavelet analysis. Brain Topogr **19**(1–2), 53–60 (2006)
15. Simos, P., Papanikolaou, E., Sakkalis, E., Micheloyannis, S. Modulation of gamma-band spectral power by cognitive task complexity. Brain Topogr **14**, 191–196 (2002)
16. Stoica, P., Moses, R. Introduction to Spectral Analysis. Prentice-Hall, Upper Saddle River, NJ (1997). 24-26

17. Torrence, C., Compo, G. A practical guide to wavelet analysis. Bull Am Meteorol Soc **79**, 61–78 (1998)
18. Weiss, S., Mueller, H. The contribution of EEG coherence to the investigation of language. Brain Lang **85**, 325–343 (2003)
19. Zar, J.: Biostatistical Analysis, 4th edn. Prentice Hall, Upper Saddle River, NJ (1999)