FEATURE EXTRACTION AND SELECTION FOR AUTOMATIC SLEEP STAGING USING EEG

Hugo Simões, Gabriel Pires
Institute of Systems and Robotics, University of Coimbra, Coimbra,Portugal
hugo@isr.uc.pt, gpires@isr.uc.pt

Urbano Nunes, Vitor Silva
Department of Electrical Engineering,University of Coimbra – Polo II, Coimbra, Portugal

Keywords: Feature extraction, feature selection, EEG sleep staging, Bayesian classifier.

Abstract: Sleep disorders affect a great percentage of the population. The diagnostic of these disorders is usually made by a polysomnography, requiring patient's hospitalization. Low cost ambulatory diagnostic devices can in certain cases be used, especially when there is no need of a full or rigorous sleep staging. In this paper, several methods to extract features from 6 EEG channels are described in order to evaluate their performance. The features are selected using the R-square Pearson correlation coefficient (Guyon and Elisseeff, 2003), providing this way a Bayesian classifier with the most discriminative features. The results demonstrate the effectiveness of the methods to discriminate several sleep stages, and ranks the several feature extraction methods. The best discrimination was achieved for relative spectral power, slow wave index, harmonic parameters and Hjorth parameters.

1 INTRODUCTION

About a third of the population suffers from sleep disorders, including the obstructive sleep apnea syndrome (Doroshenkov *et al*, 2007). The diagnosis of such diseases is performed by a polysomnography (PSG) which requires the patient's hospitalization with costs and discomfort for the patient. Ambulatory diagnostic devices may have an important role in order to mitigate these factors. The PSG consists on the acquisition of various electrical biosignals including electroencephalogram (EEG), electrooculogram (EOG) and electromyogram (EMG). The signals are segmented into epochs of 30 seconds and assigned to a sleep stage by an expert (Iber *et al*, 2007). This is a tedious and time consuming task. Automatic sleep stages classification (ASSC) is therefore an attractive solution. However, the general opinion is that most of the experts do not rely on ASSC software, because they usually present a low performance (i.e. present a high level of disagreement). One of the main reasons is due to the high variability between subjects which makes it difficult to obtain robust models for classification. The expert uses sometimes heuristics difficult to implement in the algorithms and combines a macro and micro perspective of the overall epochs. It should be highlighted that there is also some level of disagreement between experts.

This work describes part of an apnea detection system to be used in ambulatory situations by patients at home. It does not intend to substitute the PSG, but only to determine primarily if the patient is sleeping at the occurrence of the apnea episode, and secondly to determine in which sleeping stage it did occur. The stage classification relies only on EEG signals. This paper investigates several feature extraction methods to compare their performance aiming to achieve improved results in the following sleep detection stages: wake (W) vs. sleep (S), NREM (NR) sleep vs. REM (R) sleep, NREM N1 vs. NREM N2 + NREM N3, NREM N1 + NREM N2 vs. NREM N3, NREM N1 vs. NREM N2, NREM N2 vs. NREM N3 and NREM N1 vs. REM sleep (Iber *et al*, 2007). Moreover, a feature selection method based on the squared Pearson correlation coefficient (Guyon and Elisseeff, 2003), henceforth designated R-square criteria, is applied with the purpose of finding a reduced set of discriminative features. These features are used to provide additional information to the expert, and also to automatically classify each sleep stage with

some degree of certainty. The classification is performed by a Bayesian classifier using 2-class detection. Scoring sleep is done according to rules of the American Academy of Sleep Medicine (AASM) Manual for Scoring Sleep (Iber *et al*, 2007), an actualization of the rules of Rechtschaffen and Kales (Rechtschaffen and Kales, 1968). According to AASM Manual, sleep is divided into five stages: wake, NREM (Non Rapid Eye Movement) sleep (N1, N2 and N3) and REM (Rapid Eye Movement) sleep. Considering only EEG signals, the wake stage is characterized by a low amplitude alpha activity (8-13 Hz); N1 by a low amplitude theta activity (3-7 Hz); in N2 the predominant frequencies are in the 0.7-4 Hz range and there is the arising of sleep spindles and K-complexes; N3 presents at least 20% of the epochs with delta activity (<2 Hz) with amplitude greater than 75 μV; REM is characterized by frequencies mostly between 2 and 6 Hz with low amplitude. Sleep staging based only on EEG presents some difficulties because different stages such as wake, REM and NREM N1 present similar patterns. The ASSC has been addressed by many research groups. In (Tang *et al*, 2007), Hilbert-Hang transform and wavelet transform were applied to extract harmonic parameters from EEG signals, (Hese *et al*, 2001) implemented a semi-automatic method based on k-means clustering algorithm. (Ebrahimi *et al*, 2008) used neuronal networks and wavelet packet coefficients to discriminate between different sleep stages. Doroshenkov *et al*. (2007) have developed a classification algorithm based on Hidden Markov Models using only EEG signals. (Zoubek *et al*, 2007) have used feature selection algorithms to find the relevant features extracted from PSG signals. Schwaibold *et al* (2003) have implemented a neuro-fuzzy algorithm to model the rules of Rechtschaffen and Kales. Although some studies show good performance, they are very limited to specific groups of patients and it has not been possible yet to create generalized models that provide results accepted by the experts. Moreover, it remains difficult to discriminate between certain sleep stages using only EEG signals.

2 DATABASE

Data from all-night PSG records were provided by the Laboratory of Sleep from *Centro Hospitalar de Coimbra*. The PSG was recorded by the model Somnostar Pro from Viasys at a sampling frequency of 200 Hz. The database comprises seven patients (five males and two females) with ages between 27 and 64 years old (mean = 50 years; standard deviation = 12.88 years). Only six EEG channels were used: F3-A2, C3-A2, O1-A2, F4-A1, C4-A1 and O2-A1. All recordings were segmented into epochs of 30 seconds and labelled by an expert.

The dataset was initially composed by 6558 epochs. In order to avoid the over-fitting in the learning and testing of algorithms, the number of sleep epochs in the database was reduced to 3000, balancing the distribution of epochs of different sleep stages according to a normal night sleep distribution as presented in Table 1. Since the sleep stages N2 and N1 are the ones with the highest and lowest occurrence during a normal night sleep, respectively, they were set as the stages with major and minor number of epochs in the dataset, respectively, and the other sleep stages have a number of epochs between these limits.

Table 1: Full and reduced datasets.

Sleep Stages	Wake	NREM			REM
		N1	N2	N3	
Full dataset	1293	784	2431	1154	896
Reduced dataset	560	410	760	520	750

3 AUTOMATIC SLEEP SCORING

The classification methodology is illustrated in the block diagram presented in figure 1. The EEG signals are filtered and segmented. Different types of features extraction are used. These features are then selected using the correlation criteria R-square measure in order to provide the classification stage,

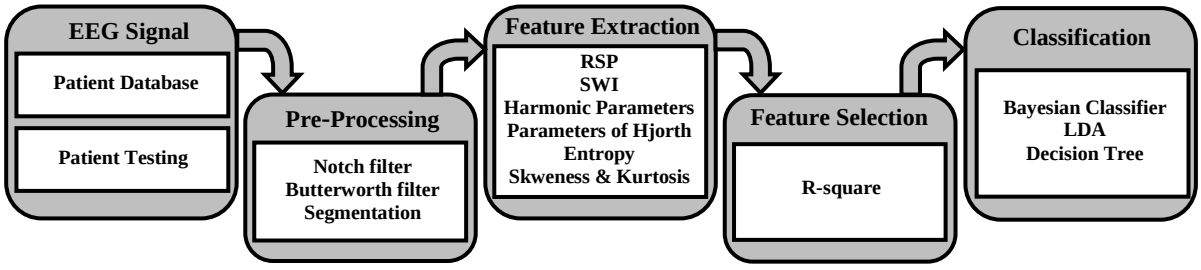

Figure 1: Classification methodology.

a Bayesian-based classifier, with the most discriminative ones. The training process uses data from a pool of patients and some data from the patient being monitored, namely, the wake recorded epochs before the patient fall asleep. This way, the wake model can be improved. Moreover, the wake epochs can be used for calibration of sleep stages. The performance analysis of the of feature extraction algorithms was done through ten-fold cross validation. The patients' database is partitioned into ten groups with the same number of epochs from each sleep stage. Nine of them are used to perform the models of classification and one for testing. This process is repeated 10 times using a different group for testing.

4 FEATURE EXTRACTION AND SELECTION

In ASSC, the EEG is traditionally analyzed in frequency domain because, according with AASM Manual, each sleep stage is essentially distinguished by some spectral properties. However, temporal analysis provides also useful information. For each EEG channel, 34 features were extracted using several methods as described in the following.

Spectral analysis provides some of the most important features. For each sleep epoch, an autoregressive method solved by the Yule-Walker algorithm was applied to estimate the power spectral density (PSD) (Yilmaz *et al*, 2007). The spectrum is divided into ten frequency sub-bands as represented in Table 2. For each sub-band, the relative spectral power (RSP) was computed. This parameter is given by the ratio between the sub-band spectral power (BSP) and the total spectral power, i.e., the sum of all 10 BSP sub-bands. This normalization is important to increase classification robustness during the recording session.

Table 2: Spectral sub-bands used in RSP computation.

Bands	**Sub-bands**	**Bandwidth $\{f_L,f_H\}$ (Hz)**
Delta	Delta 1	{0.5,2.0}
	Delta 2	{2.0,4.0}
Theta	Theta 1	{4.0,6.0}
	Theta 2	{6.0,8.0}
Alpha	Alpha 1	{8.0,10.0}
	Alpha 2	{10.0,12.0}
Sigma	Sigma 1	{12.0,14.0}
	Sigma 2	{14.0,16.0}
Beta	Beta 1	{16.0,25.0}
	Beta 2	{25.0,35.0}

Some spectral bands can be highlighted over slow wave bands by means of slow wave index (SWI) defined by the following ratios:

$$DSI = BSP_{Delta} / (BSP_{Theta} + BSP_{Alpha}) \quad (1)$$

$$TSI = BSP_{Theta} / (BSP_{Delta} + BSP_{Alpha}) \quad (2)$$

$$ASI = BSP_{Alpha} / (BSP_{Delta} + BSP_{Theta}), \quad (3)$$

where DSI, TSI and ASI stand for delta-slow-wave index, theta-slow-wave index and alpha-slow-wave index, respectively (Agarwal *et al*, 2001).

Harmonic parameters allow the analysis of a specific band in the EEG spectrum. They include three parameters: center frequency (f_c), bandwidth (f_σ) and spectral value at center frequency (S_{fc}), defined as follows (Tang *et al*, 2007):

$$f_c = \sum_{f_L}^{f_H} f P_{xx}(f) \Big/ \sum_{f_L}^{f_H} P_{xx}(f) \quad (4)$$

$$f_\sigma = \left(\sum_{f_L}^{f_H} (f - f_c)^2 P_{xx}(f) \Big/ \sum_{f_L}^{f_H} P_{xx}(f) \right)^{1/2} \quad (5)$$

$$S_{f_c} = P_{xx}(f_c), \quad (6)$$

where, $P_{xx}(f)$ denotes the PSD, which is calculated for the frequency bands $\{f_L,f_H\}$ (see Table 2).

The Hjorth parameters provide dynamic temporal information of the EEG signal. Considering the epoch x, the Hjorth parameters are computed from the variance of x, var(x), and the first and second derivatives x', x'' according to (Ansari-Asl *et al*, 2007)

$$Activity = \mathrm{var}(x) \quad (7)$$

$$Mobility = \sqrt{\mathrm{var}(x')/\mathrm{var}(x)} \quad (8)$$

$$Complexity = \sqrt{\mathrm{var}(x'') \times \mathrm{var}(x) / \mathrm{var}(x')^2}\,. \quad (9)$$

The entropy gives a measure of signal disorder and can provide relevant information in the detection of some sleep disturbs. It is computed from histogram of the EEG samples of each sleep epoch, according with (Zoubek *et al*, 2007)

$$Entropy = -\sum_{i=1}^{N} \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right), \quad (10)$$

where n is the number of samples within the sleep epoch, N is the number of bins used in computation of histogram and n_i is the number of samples within the *ith* bin.

The skewness is a measure of symmetry. The kurtosis is a measure of wether the data are peaked or flat relative to a normal distribution. Defining the *kth* order moment m_k as (Zoubek *et al*, 2007)

$$m_k = \frac{1}{n}\sum_{i=1}^{n}\left(y(i)-\bar{y}\right)^k, \tag{11}$$

where n is the number of samples of an epoch and $\bar{y}$ is the mean of these samples, the skewness and kurtosis are given by

$$skewness = m_3 / m_2 \times \sqrt{m_2} \tag{12}$$

and

$$kurtosis = m_4 / m_2 \times m_2 . \tag{13}$$

Features are usually selected by wrapper or filter methods using sequential approaches. The results from wrappers methods are dependent of the choice of the classification algorithm. Our option fell on an R-square filter approach which is independent of the classifier, based on the Pearson correlation coefficient defined as (Guyon and Elisseeff, 2003):

$$\Re = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}, \tag{14}$$

where X and Y represent two random distributions of samples, and cov and var designates covariance and variance, respectively. Considering x_i and y_i as the sample values of feature i labelled with class 1 and class 2, respectively, the value $R(i)$ for the feature i is given by:

$$R(i) = \frac{\sum_{k=1}^{m}(x_{i,k}-\bar{x_i})(y_{i,k}-\bar{y_i})}{\sqrt{\sum_{k=1}^{m}(x_{i,k}-\bar{x_i})^2\sum_{k=1}^{m}(y_{i,k}-\bar{y_i})^2}}, \tag{15}$$

where $\bar{x_i}$ and $\bar{y_i}$ represent the mean value of x_i and y_i of the m samples. The R-square, computed as $R(i)^2$, provide a level of discrimination between the two classes. High values of R-square indicate large inter-class separation and small within-class variance. The R-square provides a feature discrimination ranking.

5 BAYESIAN CLASSIFICATION

The conditional density function of the class i is modelled as a multivariate distribution under gaussian assumption

$$P(Y \mid \mu_i, \Sigma_i) = K \exp\left(-(Y-\mu_i)^T \Sigma_i^{-1}(Y-\mu_i)/2\right) \tag{16}$$

where,

$$K = 1/\left((2\pi)^{n/2}|\Sigma_i|^{1/2}\right), \tag{17}$$

Y is the feature vector resulting from concatenation of the extracted features, μ_i and Σ_i are respectively, the mean and covariance matrices computed for each class w_i from the training data. The Bayes decision function is written as:

$$\hat{w}(Y) = \arg\max\{\{\Delta_2 p(Y \mid w_1)P(w_1)\}, \{\Delta_1 p(Y \mid w_2)P(w_2)\}\}, \tag{18}$$

where $P(w_i)$ is the *ith* class prior probability and Δ_i an adjustment parameter to control the rate of false positives and false negatives (Heijden *et al*, 2004).

6 RESULTS AND DISCUSSION

The feature extraction process provides a vector of 204 features, 34 features per each EEG channel: 10 RSP, 3 SWI, 15 harmonic parameters, 3 Hjorth Parameters, 1 entropy feature, 1 skewness and 1 kurtosis. Next, the features are sorted in a decreasing order of level of discrimination by applying the R-squared based selection approach. Figure 2 shows the percentage of disagreement for wake/sleep detection between our ASSC system and expert classification (i.e. the percentage of epochs for which the automatic classification differs from manual classification made by the expert), as function of the number of features, i. e., the n-most discriminative features with n = 1,…,52. The disagreement values are obtained from a ten-fold cross validation. The lowest disagreement value was reached using the first 19 ranked features. Table 3 presents the results for each binary classifier, using 1, 2, 3, 19 most discriminative features and all 204 features. Selecting the relevant features reduces the number of features used in the ASSC leading to an increased robustness of the classifiers.

The feature selection also enables to identify the type of features and channels that lead to higher discrimination results for each 2-class discriminator

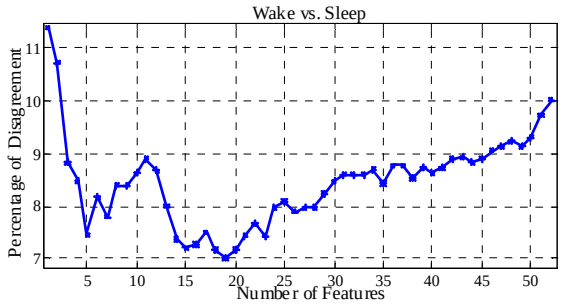

Figure 2: Percentage of disagreement vs. number of features used in wake vs. sleep classification.

Table 3: Percentage of disagreement obtained using 1, 2, 3 and the 19 most discriminative features and all 204 features.

	1	2	3	19	204
W vs. S	11,4	10,7	8,8	7,0	16,7
R vs NR	22,5	21,4	19,5	15,6	30,8
N1 vs. N2/N3	15,1	15,7	15,7	10,6	72,5
N1/N2 vs. N3	15,7	14,7	14,6	15,5	30,3
N1 vs. N2	21,9	22,6	18,5	15,6	63,9
N2 vs. N3	19,0	18,2	16,7	17,7	39,8
N1 vs. R	25,5	24,7	24,4	25,0	64,7
Mean	18,7	18,3	16,9	15,3	45,5

(Table 4). As it can be seen, the feature entropy (Ent), Skewness (Skw) and kurtosis (Krt) never appear in the 20 most discriminative features. On the other hand, the most frequents are the RSP and harmonic parameters. Analyzing the origin of the 20 most discriminative features for each case, the parameters of Hjorth (PHj) are most evident in N1/N2 vs. N3 and N2 vs. N3, but they have no weight in R vs. NR and N1 vs. R. The harmonic parameters are more frequent in W vs. S, N1 vs. N2/N3 and N1 vs. N2, but are not relevant in R vs. NR, N1 vs. N2/N3, N2 vs. N3 and N1 vs. R. For the RSP and SWI, they have a similar number of features in all discriminations, except for N1 vs. R, where the RSP has several features with good discrimination, and for N1 vs. N2, where SWI does not assume any importance. Analyzing the EEG channels, it can be seen that O1A2 (O1) and O2A1 (O2) are the most relevant in discrimination wake vs. sleep; F3A2 (F3) and F4A1 (F4) in REM vs. NREM; and C3A2 (C3) and C4A1 (C4) in N2 vs. N3. In the remaining discriminations, they all have a relatively uniform distribution, except in N1 vs. R, in which the channels O1A2 and O2A1 do not have any type of contribution. Figure 3 shows the type of features and channels that lead to higher discrimination results, taking all discriminators together. Summarizing, the best ranked discriminative features never include entropy features, skewness or kurtosis. These parameters are related to the signal shape. However, since the EEG signal patterns are very random, it is difficult to obtain useful information from these parameters.

Instead, the set of most discriminatory features between sleep stages was composed mainly by RSP and Harmonic Parameters. This result emphasizes the fact that the spectral analysis has more discriminative information than temporal signal analysis as already concluded in (Hese *et al*, 2001; Tang *et al*, 2007).

Table 4: Number of feature type and channels within the 20 most discriminative features.

		W vs. S	R vs. NR	N1 vs. N2/N3	N1/N2 vs. N3	N1 vs. N2	N2 vs. N3	N1 vs. R	Total
Features	**RSP**	5	4	6	6	5	6	13	45
	SWI	3	2	2	2	0	4	4	17
	HP	9	14	8	3	12	2	3	51
	PHj	3	0	4	9	3	8	0	27
	Ent	0	0	0	0	0	0	0	0
	Skw	0	0	0	0	0	0	0	0
	Krt	0	0	0	0	0	0	0	0
Channels	**F3**	1	6	6	3	4	2	5	27
	C3	1	3	4	5	4	5	5	27
	O1	6	1	3	4	2	4	0	20
	F4	2	5	3	2	4	1	5	22
	C4	5	3	3	4	4	5	5	29
	O2	5	2	1	2	2	3	0	15

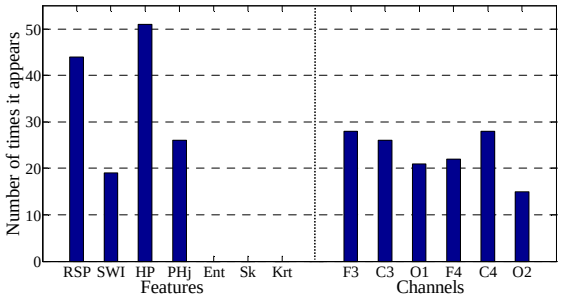

Figure 3: Number of times that each group of features and each channel appears in the 20 most discriminative features.

On the other hand, all the 6-six EEG channels provide useful features for sleep staging discrimination. Analyzing the results for each of the binary classifiers, there is greater disagreement in the case of N1 vs. R sleep. This situation relates to the fact that, in terms of EEG, the patterns presented in these two stages are very similar. Finally, a

decision tree was implemented based on 2-class detection, as represented in Figure 4. At each step, a new level was introduced from a wake/sleep to all stages classification. The results were compared with and without feature selection (Table 5). The improvements from feature selection are evident. The results obtained with our ASSC system are comparable to the ones obtained in other methods based on EEG only described in literature (zoubek *et al*, 2007; Doroshenkov *et al*, 2007).

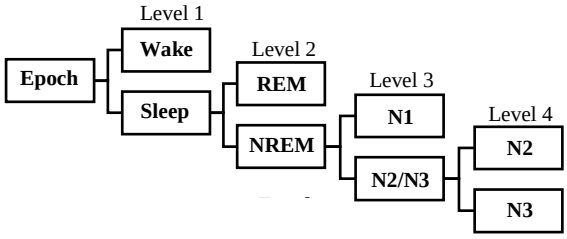

Figure 4: Decision tree based on 2-class detection.

Table 5: Disagreement obtained with using 19 most discriminative features and all 204 in 2, 3, 4 and 5 sleep stages classification.

Classification	Diasagreement (%)	
	All Features	19
2 Class	36	7
3 Class	62	18
4 Class	83	22
5 Class	83	29

7 CONCLUSIONS

In this paper, the use of several feature extraction methods was investigated in the context of EEG-based sleep staging. The first conclusion was that the most discriminative features were determined by RSP, SWI, Harmonic Parameters and Parameters of Hjorth. All the 6-EEG channels provide useful information. On the other hand, the application of the feature selection method improved, in general, the process of discrimination by selecting the set of features that provided a lower percentage of disagreement. One of the biggest problems in automatic sleep staging based on EEG is the similarity between patterns of different sleep stages such as REM and NREM N1. This can be improved recurring to other biosignals, such as EOG and EMG. Another problem in ASSC is the high level of variability between patients. Using an ambulatory system, the patient can perform periodic recordings at home. This way, the first session can be fully analysed by the expert. The labelled data can be used to obtain classification models specific to the patient. Further sessions can then use these robust user-dependent models. This approach is under research presently.

REFERENCES

Ansari-Asl. K., Chanel, G., Pun, T., A channel Selection Method for EEG Classification in Emotion Assessment Based on Synchronization Likelihood. In *EUSIPCO'07*, 1241-1245.

Doroshenkov, L., Konyshev, V., Selishchev, S., 2007, Classification of Human Sleep Stages Based on EEG Processing Using Hidden Markov Models. *Biomedical Engineering*, 41(1), 25-28.

Ebrahimi, F., Mikaeili, M., Estrada, E., Nazeran, H., 2008, Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients. In *IEEE EMBS'08*, 1, 1151-1154.

Guyon, I., Elisseeff, A., 2003, An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.

Heijden, F., Duin, R., Ridder, D., Tax, D., 2004, *Classification, Parameter Estimation and State Estimation.* John Wiley & Sons.

Hese, P., Philips, W., Koninck, J., Walle, R., Lemahieu, I., 2001, Automatic Detection of Sleep Stages Using the EEG. In *IEEE EMBS'01*,*Proc. of*, 1994-1947.

Iber., C., Ancoli-Israel, S., Chesson, A., Quan, S., 2007, *The AASM Manual for the scoring of Sleep and Associated Events: Rules. Terminology and Technical Specifications* (1st ed.). Westchester, Illinois: American Academy of Sleep Medicine.

Rechtscheffen, A., Kales, A., 1968, *A Manual of Standardized Terminology, Techniques ans Scoring System for Sleep Stages of Human Subjects*, US Government Printing Office, National Institute of Health Publications, Washington DC.

Schwaibold, M., Harms, R., Scholer, B., Pinnow, I., Cassel, W., Penzel, T., Becker, H., Bolz, A., 2003, Knowledge-Based Automatic Sleep-Stage Recognition – Reduction in the Interpretation Variability. *Somnologie*, 7, 59-65.

Tang, W. C., Lu, S. W., Tsai. X. M., Kao. C. Y., Lee. H. H., 2007, Harmonic Parameters with HHT and Wavelet Transform for Automatic Sleep Stages Scoring. *Proc. of World Academy of Science. Engineering and Technology*, 22, 414-417.

Yilmaz, A., Alkan, A., Asyali, M., 2007, Applications of parametric spectral estimation methods on detection of power system harmonics. *Electric Power Systems Research* (2008), 78, 683-693.

Zoubek, L., Charbonnier, S., Lesecq, S., Buguet, A., Chapolot, F., 2007, Feature selection for sleep/wake stages classification using data driven methods. *Biomedical Signal Processing and Control*, 2, 171-179.