A channel selection method for EEG classification in emotion assessment based on synchronization likelihood

Ansari Asl, Karim; Chanel, Guillaume; Pun, Thierry

How to cite

ANSARI ASL, Karim, CHANEL, Guillaume, PUN, Thierry. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In: 15th European Signal Processing Conference (Eusipco 2007). Poznan, Poland. [s.l.] : [s.n.], 2007. p. 1241–1245.

This publication URL: https://archive-ouverte.unige.ch/unige:21699

A CHANNEL SELECTION METHOD FOR EEG CLASSIFICATION IN EMOTION ASSESSMENT BASED ON SYNCHRONIZATION LIKELIHOOD

Karim Ansari-Asl, Guillaume Chanel, and Thierry Pun

Computer Vision and Multimedia Laboratory, Computer Science Department, University of Geneva
7 route de Drize, 1227 Carouge-Geneva, Switzerland
phone: +41 (0)223790152, email: {karim.ansari, guillaume.chanel, thierry.pun}@cui.unige.ch
web: http://vision.unige.ch

ABSTRACT

When assessing human emotion using EEG classification, one of the critical problems is to deal with the very large number of features to be classified. In this paper, we address this problem using synchronization likelihood as a new channel selection method. Applying this method, we could significantly reduce the number of EEG channels to be used in emotion assessment, with only slight (if any) loss of classification performance depending on the used feature. We report and compare the results obtained by employing a linear classifier on different features extracted either from all channels or from the selected subset of channels. These features include synchronization likelihood, Hjorth parameters, and fractal dimension.

1. INTRODUTION

Emotion assessment has recently attracted the attention of many researchers from different fields. This assessment is important from the theoretical and applied emotion research points of view. Concerning the theoretical aspect, this assessment can help to better understand the emotion mechanisms, because there is currently consensus neither on the mechanisms involved nor on the definition of emotion itself [1]. Regarding applications, emotion assessment can lead to the integration of emotions into human-machine interaction (HCI) systems. This integration will improve the quality of life, for instance for disabled and elderly people, by bringing HCI closer to human-human interaction.

Physical (or behavioral) and physiological measurements could be used for emotion assessment. Amongst them, we can cite measures related to facial expression, vocal intonation, heart rate, temperature, and electrodermal response [1]. It is well established that, based on the cognitive theory of emotion [2], brain activity plays a central role in emotion. However, few works have used it to assess the emotional state. The usage of electroencephalographic (EEG) signals in emotion assessment is reinforced by the emergence of an increasing number of practical brain-computer interfaces (BCIs), in which EEG signals are used to decipher state, thought, and decision of user.

Works that try to assess emotions using EEG signals differ in the way emotions are elicited, which has a strong impact when dealing with brain activity; since neuronal structures will not process information from different stimuli in the same way. In [3], participants are stimulated using film clips, while in [4] images from the IAPS (International Affective Picture System) are used. The current work aims at distinguishing between three emotions, exciting-positive, exciting-negative and calm-neutral, in an emotional recall paradigm. Emotion elicitation from recall of past events is used in psychological studies and has shown to activate different parts of the brain; it has also been used for emotional assessment from peripheral physiological signals [5]. To our knowledge, such an elicitation technique has never been used for emotion assessment using EEG signals.

For EEG classification various features have been used such as wavelet coefficients, autoregressive model parameters, signal energy in different frequency bands, fractal dimension, and Lyapunov exponents [6]. In dealing with EEG classification, an important problem is the huge number of features. It comes from the fact that (*i*) EEG signals are non-stationary, thus features must be computed in a time-varying manner and (*ii*) the number of EEG channels is large (for instance, 64 in our own experimental setting). Solutions to alleviate this problem, “the curse of dimensionality,” consist of feature selection and channel selection methods.

In this work, we propose a new channel selection method to choose the EEG channels which better convey the information about emotional states. Some methods have been recently proposed to tackle this problem; most of them used a modified feature selection method wherein all features from a channel are considered as a unique entity [7, 8]; only few studies tried to reduce the number of channel independently of the classification scheme [9].

Activity from each brain source projects into an area of the scalp where data is recorded by a number of EEG channels; it also propagates from one brain region to another. Consequently, we can find common information, simultaneously or with some time delay, on different EEG channels. Using this fact, one can apply relationship measurements to choose one representative among many EEG channels. Considering the number of channels, it is very difficult to deal with them by pairwise relationship measurements [10]. Thus, we opted for the Synchronization Likelihood (SL) method as a multichannel measurement [11].

Applying SL, we were able to reduce the number of EEG channels from 64 to 5. To evaluate the performance of SL in channel reduction, we have used different kind of features (SL, Hjorth parameters, and fractal dimension) and

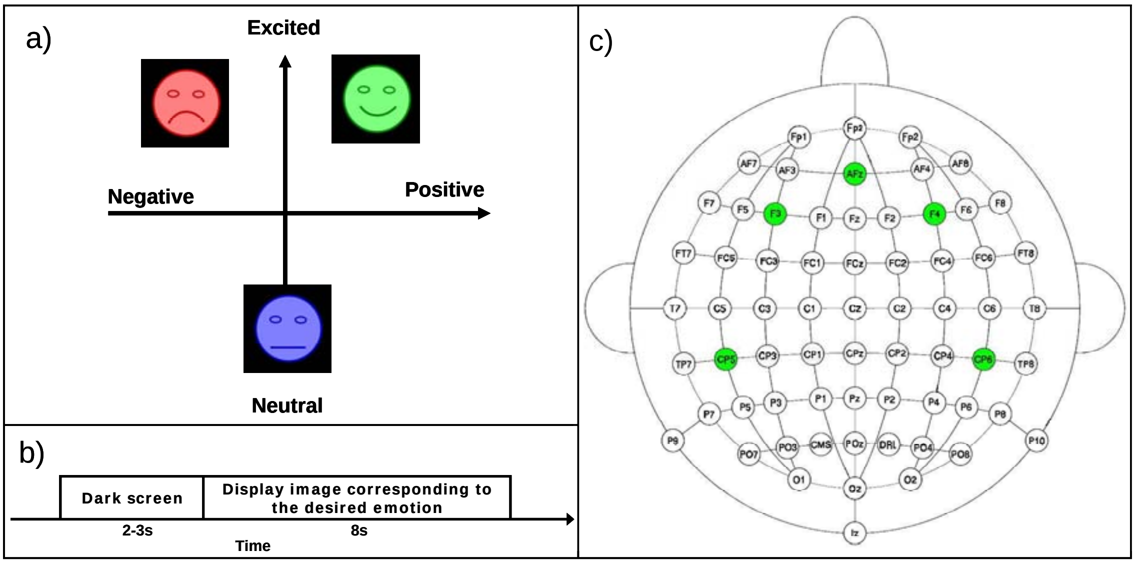

Figure 1 – (a) 2D emotions representation in the valence-arousal space, with indication of the three emotion classes considered in this study together with the visual symbol used to prime the participants. (b) Chronology of one trial in the emotional recall protocol. (c) Schematic representation of electrodes positions on the scalp for the Biosemi Active Two with 64 channel EEG and two references (CMS and DRL); the five channels selected by SL as the representative set are marked in green color.

different classifiers; to be short we report here only the results of a linear classifier. Results showed that the channel reduction leads only to a slight (if any) loss of classification accuracy. Since a final goal is to integrate emotions into a real-time human-machine interaction system, this slight loss is negligible in relation to the reduction in computation time.

2. PROTOCOL AND DATA ACQUISITION

In order to acquire physiological data under different emotional states, we designed a protocol based on the recall of strong emotional events. Before the beginning of the recording, participants are asked to retrieve two emotional events from their past life, one exciting-negative (class one, C1) and one exciting-positive (class three, C3). To define a third state referred to as “calm-neutral,” the participant is asked to stay calm and relax (class two, C2).

The participant sits in front of a computer displaying the images to inform him/her about the specific emotional event he/she has to think of. Fig. 1(a) shows the locations, in the valence-arousal space, of the three images corresponding to different emotional recalls. Prior to displaying each image, a dark screen is shown for 2 to 3 seconds to separate each trial; there is also a rest period of undetermined duration after each 30 trials. This protocol enables us to elicit emotions corresponding to the three classes; only these three regions of the valence-arousal space are used since there are very few emotions that are either calm-negative or calm-positive. Fig. 1(b) shows the chronology of one trial. We recorded 100 trials for each emotional class, *i.e.* in total 300 trials.

For the moment, only one male participant (right handed, 30 years old) took part in this study. We are recording the emotional activity of other participants to validate the current results (work in progress). EEG signals are recorded using the Biosemi Active Two device (www.biosemi.com) comprising 64 electrodes (plus 2 for reference, namely CMS and DRL electrodes); the positions of electrodes are presented in Fig. 1(c). Based on reports from the literature, signals were sampled at 1024 Hz and filtered between 4-45 Hz in the preprocessing step.

3. SYNCHRONIZATION LIKELIHOOD

Synchronization likelihood is a measure of generalized synchronization in multivariate data sets [11]; it measures both the linear and nonlinear interdependency between signals. Given *M* time series $x_{k,n}$, where *k* denotes the channel number ($k = 1,\ldots,M$) and *n* stands for discrete time samples ($n = 1,\ldots,N$), the state space corresponding to each time series is reconstructed using the time-delayed embedding method [12]:

$$X_{k,n} = \left(x_{k,n}, x_{k,n+\tau}, x_{k,n+2\tau}, \ldots, x_{k,n+(d-1)\tau} \right),$$

where τ is the time lag and *d* is the embedding dimension.

The probability that embedded vectors $X_{k,n}$ and $X_{k,m}$ ($w_1 < |n-m| < w_2$) are closer to each other than a distance ε is:

$$P_{k,n}^{\varepsilon} = \frac{1}{2(w_2 - w_1)} \sum_{\substack{m=1 \\ w_1 < |n-m| < w_2}}^{N} \theta\left(\varepsilon - \left| X_{k,n} - X_{k,m} \right|\right),$$

where $|\cdot|$ denotes Euclidean distance, θ stands for Heaviside step function, w_1 is the Theiler correction [13], and w_2 determines the length of sliding window, i.e. the time resolution of the synchronization measure. Letting $P_{k,n}^{\varepsilon} = P_{ref} \ll 1$ be a small arbitrary probability, the above equation for $X_{k,n}$ gives a critical distance $\varepsilon_{k,n}$ as the neighbourhood criterion. We can now determine for each discrete time pair (*n*, *m*), within our considered window ($w_1 < |n-m| < w_2$), the number of channels $H_{n,m}$ whose embedded vectors $X_{k,n}$ and $X_{k,m}$ are closer than $\varepsilon_{k,n}$:

$$H_{n,m} = \sum_{k=1}^{M} \theta\left(\varepsilon_{k,n} - \left|X_{k,n} - X_{k,m}\right|\right).$$

The synchronization likelihood for each channel *k* and each discrete time pair (*n*, *m*) is defined as follows:

$$S_{k,n,m} = \begin{cases} \dfrac{H_{n,m}-1}{M-1} & if \ \left|X_{k,n} - X_{k,m}\right| < \varepsilon_{k,n} \\ 0 & if \ \left|X_{k,n} - X_{k,m}\right| \geq \varepsilon_{k,n} \end{cases}.$$

Finally, by averaging over all discrete times *m*, the synchronization likelihood for channel *k* at discrete time *n* can be obtained:

$$S_{k,n} = \frac{1}{2P_{ref}\left(w_2 - w_1\right)} \sum_{\substack{m=1 \\ w_1 < |n-m| < w_2}}^{N} S_{k,n,m}.$$

Synchronization likelihood, $P_{ref} \leq S_{k,n} \leq 1$, determines how strongly channel *k* at time *n* is synchronized with all other *M*-1 channels; $S_{k,n} = P_{ref}$ corresponds to the completely uncorrelated case, whereas $S_{k,n} = 1$ corresponds to the completely synchronized case.

4. FEATURE EXTRACTION

We have used several different features to discriminate the EEG signal recorded during different emotional states. These features include synchronization likelihood, Hjorth parameters (activity, mobility, and complexity), and fractal dimension.

4.1 Synchronization Likelihood

Synchronization measures have been proved useful in discriminating EEG signals in a BCI framework [10]. In the current work, we used synchronization likelihood as a feature to discriminate EEG signals; its definition has been reviewed above.

4.2 Hjorth Parameters

Three Hjorth parameters [14] for a signal *x* of length *N*, namely *activity*, *mobility* and *complexity*, are defined below. Activity is equal to the variance $\mathrm{var}(x)$ of the signal *x*:

$$Activity(x) = \frac{\sum_{n=1}^{N}\left(x(n) - \bar{x}\right)^2}{N},$$

where $\bar{x}$ denotes the mean of *x*. Mobility is a measure of the signal mean frequency:

$$Mobility(x) = \sqrt{\frac{\mathrm{var}(x')}{\mathrm{var}(x)}},$$

where x' stands for the derivate of signal *x*. Complexity measures the deviation of the signal from the sine shape:

$$Complexity(x) = \frac{Mobility(x')}{Mobility(x)}.$$

4.3 Fractal Dimension

The fractal dimension (FD) is a quantity that conveys information about the space filling and self-similarity of an object. For a time series, as a waveform, it is natural to expect a FD between one (FD of a straight line) and two (FD of a plane).

A comparative study of different waveform FD computing methods showed that, in the application of FD to EEG signals, Katz method outperforms others [15]. Here, we used a method from a more recent study [16] reporting better performance than Katz method; it computes directly FD from the waveform. At the first step, the time series x with length *N* is normalized with respect to time and amplitude:

$$n^* = \frac{n}{N},$$

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum of the signal amplitude, respectively. Now the FD of waveform can be approximated as follows:

$$D = 1 + \frac{\ln(L)}{\ln(2M)},$$

where L is the length of the normalized curve, and $M = N - 1$. *L* is computed as the sum of Euclidean distances of all adjacent pairs of waveform points:

$$L = \sum_{n=2}^{N} \sqrt{\left(x^*(n) - x^*(n-1)\right)^2 + \left(n^* - (n-1)^*\right)^2}.$$

5. CLASSIFIER

To be short, in this paper we only report the results of a simple linear classifier, i.e. diagonal Linear Discriminant Analysis (LDA) [17]. With this classifier a multivariate normal density distribution is fitted on each group. In the diagonalized version of linear discriminant analysis, each distribution is assumed to have the same diagonal covariance matrix. Means and variances of distributions are computed from learning samples of the database and classification is then applied to a test sample; this step is repeated several times using a leave one-out strategy to obtain the mean accuracy of the classifier.

6. RESULTS

We applied the synchronization likelihood approach to EEG signals recorded during the emotional recall of three classes of emotions, exiting-positive, exiting-negative, and neutral. The first step to compute the SL consisted of the determination of the parameters of the reconstructed state space for the EEG signals, i.e. *d* and τ. Its dimension, *d*, was computed using a modified false nearest neighbors method, proposed by Cao [18]; the lag shift, τ, was taken empirically equal to the time of first zero crossing of the autocorrelation function. In addition, the Theiler correction (w_1) was chosen to be equal to τ. For the EEG signals used in this paper, we obtained $d = 15$ and $\tau = w_1 = 30$ sample points.

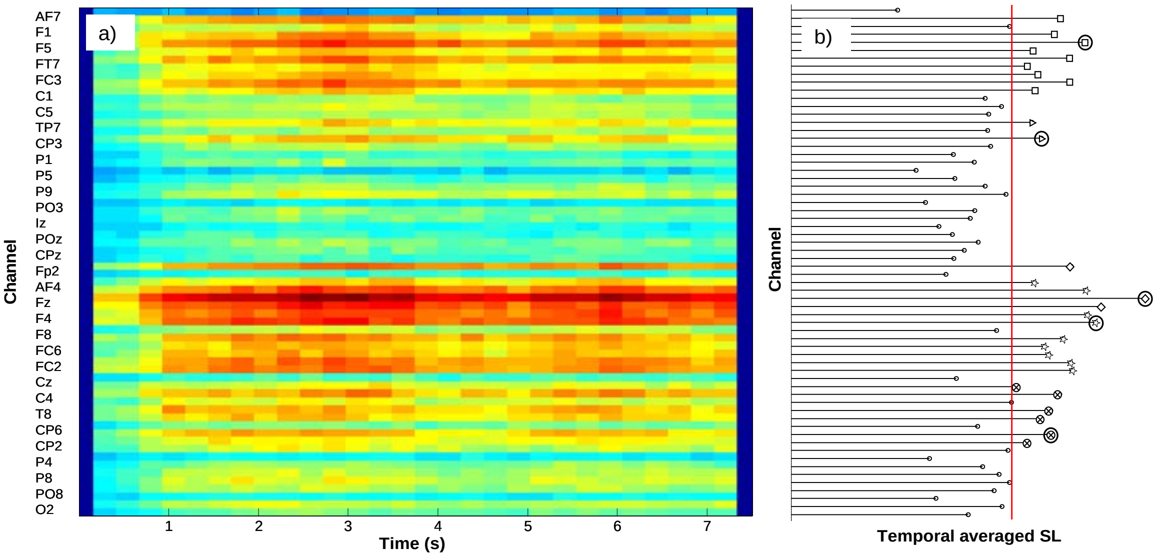

Figure 2 – Mean of SL for EEG signals corresponding to class one trials. (a) Temporal evolution of SL for all recorded channels (red and blue colors respectively correspond to the most and least significant SL values). (b) Temporal average of SL values presented in part (a) for all channels. The red line corresponds to the global mean of SL value. Channels which exceed the global mean of SL value are marked by: □, △, and ◇ for channels from the left side, center, and the right side of the anterior scalp; ☆ and ⊗ for the left and right sides of the central-posterior scalp. Selected channels from each scalp area are encircled.

Fig. 2 shows the mean SL for EEG signals corresponding to the class one trials; here SL was computed in a sliding window with a length of 512 sample points and the displacement step of 256 sample points. Fig. 2(a) shows the time evolution of SL for all channels; the average of this figure over time axis is shown in Fig. 2(b). Different scenarios have been experimented to choose the best channels as the representative set for all channels. Finally, we opted for taking into account both SL values and anatomical knowledge.

Five channels (marked by green color in Fig. 1 (c) and encircled in Fig. 2(b)), from five different brain regions, have been chosen among the EEG channels which exceed the global mean of SL value (red line in Fig. 2(b)). Three of them are from the anterior part of the scalp: F3, F4, and AFz which have maximal SL values among channels from the left side, the right side, and the center (marked by □, △, and ◇ in Fig. 2(b)). Among the channels located in the central-posterior part of the scalp, from the left side (marked by ☆ in Fig. 2(b)) CP5 has the maximal SL value and has been selected; from the right side (marked by ⊗ in Fig. 2(b)) we have chosen CP6, even if it does not have the maximal SL value. Since CP6 is in symmetry with CP5 and has been used as pair of CP5 in emotions studies [19], we compared it with CP2 which has the maximal SL value and found that CP6 leads to better classification performance.

The three selected channels from the anterior part of the scalp belong to the emotional circuit, because the implication of the orbito-frontal cortex and amygdala in emotion processes has been already underscored [20]. The two other selected channels from the central-posterior part of the scalp may have some relationship to the memory process in our emotional recall protocol, considering the implications of parietal and temporal lobes in memory. Besides, we observe more synchronization in the right hemisphere than in the left one; this may corroborate the well-known involvement of the right hemisphere in the non-verbal memory processes [21]. The results of applying SL to the other two classes also led to the same result as of the first one.

Table 1 and 2 respectively present the results of the LDA classifier for all EEG channels and for the five selected channels using different features. In these tables the first row corresponds to the classification of three classes, the last row before the Mean row corresponds to the classification of class two (i.e. "calm-neutral" state) versus two other classes, and the other rows correspond to the pairwise classification between classes. All features, based on a local stationarity hypothesis, were computed in sliding windows of length 512 sample points with a displacement step of 256 sample points. As can be seen from these tables, by selecting the best channel representative set we observe for the case of SL feature, that the classification rates have been increased by 5.2% in average accuracy. For the other features, we observe a slight decrease of performance, always less than 4.8% in average accuracy. In these tables, the last row and column, named Mean, are respectively the average value of other rows and columns. The global loss mean for all classes and all features is only 1.6%.

This slight loss of classification rate is well compensated by a considerable reduction in computation time. By reducing the number of channels from 64 to 5, the computation time for classification is at least divided by ten as the complexity is linear with respect to the number of channels, *e.g.* in the case of the Hjorth feature, the classification takes in average 397 and 35 second respectively for all and for five selected channels, using MATLAB® on a computer with a Pentium4 (2 GHz) processor. This reduction is of importance in view of the fact that emotion recognition algorithms should be integrated in practical HCI systems.

Table 1 – Classification accuracy (%) of LDA for all channels. Used features are SL, activity (Act.), mobility (Mob.), complexity (Com.), Hjorth as concatenation of three preceding features, and FD.

Classes	SL	Act.	Mob.	Com.	Hjorth	FD	Mean
C_1, C_2, C_3	49.3	42	50.3	45.7	51.7	49.3	48.6
C_1, C_2	58	60.5	63	61	63	62.5	61.3
C_2, C_3	68	61.5	66.5	60	67.5	67.5	65.2
C_1, C_3	66	59.5	66	58.5	67.5	63	63.4
C_2, [C_1, C_3]	59.7	58	63.7	60	63.3	64.3	61.5
Mean	60.2	56.7	61.9	57	62.6	61.3	60

Table 2 – Classification accuracy (%) of LDA for five selected channels. Used features are SL, activity (Act.), mobility (Mob.), complexity (Com.), Hjorth as concatenation of three preceding features, and FD.

Classes	SL	Act.	Mob.	Com.	Hjorth	FD	Mean
C_1, C_2, C_3	55.3	39.7	43.7	43	48	48	46.3
C_1, C_2	64.5	57.5	50.5	61	61	53.5	58
C_2, C_3	74	57	67.5	56.5	64.5	66	64.3
C_1, C_3	64.5	56	66.5	56.5	64.5	68	62.7
C_2, [C_1, C_3]	68.7	56.7	57.3	59.3	60.3	62.7	60.8
Mean	65.4	53.4	57.1	55.2	59.7	59.6	58.4

7. CONCLUSION

Emotion assessment is of importance from different point of views, such as emotion integration into HCI systems. Different physical and physiological cues including brain activity are used for emotion assessment. In EEG classification, one of the major problems is the huge number of features to be classified.

In the present paper, we report the result of the application of synchronization likelihood as a channel selection method to alleviate “the curse of dimensionality” in EEG classification. Using SL, we could reduce the number of EEG channels from 64 to 5 for EEG classification in emotion assessment. To evaluate the performance of SL in channel reduction, we have used different kind of features (SL, Hjorth parameters, and fractal dimension) and different classifiers (such as the LDA, the linear support vector machine [SVM], and the SVM using radial basis functions [RBF] as kernel); to be short we report here only the results of a linear classifier, LDA. Results showed that the channel reduction leads only to a slight (if any) loss of classification accuracy rate. Reducing the number of channels contributes to a significant decrease in computation time for emotion assessment. This decrease is of great interest for systems with low computational capabilities such as ambulatory real-time HCI systems.

8. AKNOWLEDGMENT

This work was supported in part by the EU network of excellence SIMILAR (http://www.similar.cc).

REFERENCES

[1] R. W. Picard, *Affective computing*. Cambridge, Mass.: MIT Press, 1997.

[2] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks*, vol. 18, pp. 317-52, 2005.

[3] K. Takahashi, "Remarks on Emotion Recognition from Bio-Potential Signals," in *Proc. ICARA 2004*, Palmerston North, New Zealand, Dec. 13-15. 2004, pp. 186-191.

[4] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," in *Proc. MRCS 2006*, Istanbul, Turkey, Sept. 11-13, 2006, pp. 530-537.

[5] B. Herbelin, P. Benzaki, F. Riquier, O. Renault, and D. Thalmann, "Using physiological measures for emotional assessment: a computer-aided tool for cognitive and behavioural therapy," in *Proc. 5th Inter. Conf. on Disability*, Oxford, UK, 2004, pp. 307-314.

[6] R. Boostani and M. H. Moradi, "A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier," *J. Neural Eng.*, vol. 1, pp. 212-7, 2004.

[7] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1003-10, 2004.

[8] L. Tian, D. Erdogmus, A. Adami, M. Pavel, and S. Mathan, "Salient EEG Channel Selection in Brain Computer Interfaces by Mutual Information Maximization," in *Proc. EMBC 2005*, Shanghai, China, Sept. 1-4. 2005, pp. 7064- 7067.

[9] O. Kok-Meng, K. H. Thung, W. Chong-Yaw, and R. Paramesran,"Selection of a Subset of EEG Channels using PCA to classify Alcoholics and Non-alcoholics," in *Proc. EMBC 2005*, Shanghai, China, Sept. 1-4. 2005, pp. 4195-4198.

[10] E. Gysels and P. Celka, "Phase synchronization for the recognition of mental tasks in a brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, pp. 406-15, 2004.

[11] C. J. Stam and B. W. van Dijk, "Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets," *Physica D: Nonlinear Phenomena*, vol. 163, pp. 236-251, 2002.

[12] F. Takens, "Lecture Notes in Mathematics," *Springer*, vol. 898, pp. 366, 1981.

[13] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical Review A*, vol. 34, pp. 2427-2432, 1986.

[14] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, pp. 306-310, 1970.

[15] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt,"A comparison of waveform fractal dimension algorithms," *IEEE Trans. Circuits and Systems*, vol. 48, pp. 177-183, 2001.

[16] C. Sevcik, "On fractal dimension of waveforms," *Chaos, Solitons & Fractals*, vol. 28, pp. 579-580, 2006.

[17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. New York: Wiley, 2001.

[18] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, pp. 43-50, 1997.

[19] A. Holmes, J. S. Winston, and M. Eimer, "The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression," *Brain Res Cogn Brain Res*, vol. 25, pp. 508-20, 2005.

[20] J. G. Taylor and N. F. Fragopanagos, "The interaction of attention and emotion," *Neural Networks*, vol. 18, pp. 353-369, 2005.

[21] E. A. Maguire and L. Cipolotti, "Selective sparing of topographical memory," *J. Neurol Neurosurg. Psychiatry*, vol. 65, pp. 903-9, 1998.