

A NOVEL SPACE-TIME-FREQUENCY MASKING APPROACH FOR QUANTIFICATION OF EEG SOURCE PROPAGATION WITH AN APPLICATION TO BRAIN COMPUTER INTERFACING

Leor Shoker[‡], Saeid Sanei[‡], Kianoush Nazarpour[‡], and Alex Sumich[†]

[‡] Centre of Digital Signal Processing, Cardiff School of Engineering,
Cardiff University, Cardiff, CF24 3AA, UK

[†] Brain Image Analysis Unit, Institute of Psychiatry, London, SE5 8AF
e-mails: [‡]{shokerl, saneis, nazarpour}@cf.ac.uk, [†]a.sumich@iop.kcl.ac.uk

ABSTRACT

A robust space-time-frequency signal extraction algorithm has been developed with an application to brain computer interface (BCI). The algorithm is based on extending time-frequency masking methods to accommodate the spatial domain. The space-time-frequency masks are then clustered in order to extract the desired source. Then the motion of the extracted source is tracked over the scalp. Finally, the trials are classified based on their directionality and locations over the scalp. The proposed method outperforms traditional systems by exploiting the motion of the sources.

1. INTRODUCTION

A brain computer interface (BCI) is a system which allows the user to interact with a computer using brain signals only. BCIs can be divided into two main categories; invasive and non-invasive. The former uses intracranial electrodes or subdural implanted deep inside or on the surface of the brain, whereas the latter uses surface electrodes placed over the scalp. Here, we will focus on non-invasive BCIs. Current BCIs use one of a number of extractable EEG signals, such as rhythmicities [1] in the data or a particular component, such as slow cortical potentials (SCP) [2], or evoked potentials (EPs) [3]. EPs such as P300 are time-locked events which are, generally, extracted by averaging many trials of the same event.

The authors in [4] demonstrated that there is a causal relationship between spatially neighboring channels of the EEGs. Further works in [5] [6] showed that this can be used to distinguish between left and right finger movements from the EEG.

In this paper we demonstrate that in addition to the time and frequency information of the EEG signals, the spatial and directional information provide crucial indicators of intended left or right finger movement. A block diagram of the proposed system is shown in Fig. 1. In the first section the EEGs are converted into the time-frequency (TF) domain, then the TF representation of each electrode is arranged into a matrix where each element represents the x-y coordinates of the electrode. In the next block a space-time-frequency mask is created and the components within the mask are clustered. The cluster centres are one of the features used by the classifier. The other significant feature is the directionality of the moving reconstructed source signal which is deduced from its cross correlation with the raw EEGs.

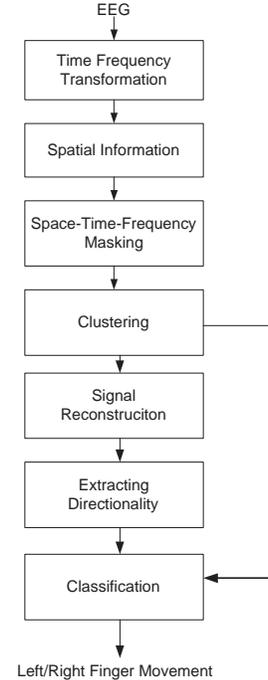


Figure 1: Block diagram of space-time-frequency based atom extraction and classification algorithm.

2. METHODS

In this section we show that cortical regions can be separated by assuming that the EEG sources are disjoint in space, time, and frequency. Section 2.1 explains the method for extracting the space-time-frequency distribution (STFD) from the EEGs. Then section 2.2 describes the clustering technique for extracting the atoms from the space-time-frequency distributions. Section 2.3 explains the reconstruction of the signals from the clustered STFD. In section 2.4 we describe the motion characterisation algorithm which forms one of the features used in the classification algorithm described in section 2.5.

2.1 Space-Time-Frequency Analysis

The time frequency distribution (TFD) of each electrode is constructed using short term Fourier Transform (STFT) de-

fined by

$$F_{s_j}(t, f) = \frac{1}{\sqrt{2\pi}} \sum_{\tau} w(t - \tau) s_j(t) e^{-i\omega\tau} \quad (1)$$

where $w(\cdot)$ is a window function and $s_j(t)$ is the j -th electrode. The time-frequency plot for each electrode is arranged into a four dimensional matrix such as

$$P(x, y, t, f) \triangleq F_{s_j} \quad j = 1, \dots, N \quad (2)$$

where x and y are the spatial coordinations of the electrodes, t is the time index, and f is the frequency index. The first two dimensions, x and y are sufficiently large so that all the electrodes can be arranged as they are defined by the 10-20 electrode placement system. For example, 11×11 matrix is sufficient for a 64 electrode EEG. The electrode Cz would be located at coordinates (6,6). The parameter N is the number of electrodes.

A space-time-frequency mask is constructed from the STFD plot based on the following criterion,

$$M(x, y, t, f) = \begin{cases} 1, & 20 \log(P(x, y, t, f)) > u \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where u is a threshold, which is empirically chosen to be $0.25 \max(F_{s_j})$.

2.2 Clustering

In order to extract the atoms from the STFD, the regions of activity (atoms) in space-time-frequency must be identified and isolated from the background EEG. We used the kmean clustering algorithm [7] to identify and separate the active regions in the STFD mask $M(x, y, t, f)$. Since the number of clusters, k , is unknown we first estimate the optimum number of clusters by using the GAP statistic method [8]. The cluster compactness is given by

$$h_k = \sum_{r=1}^k \sum_{i \in C_r} \|\mathbf{q}_i^r - \mathbf{m}_r\|^2 \quad (4)$$

where \mathbf{q}_i^r are the points within the cluster, $k \in \{1, \dots, K\}$ is the number of clusters and K is the maximum number of clusters, C_r are points within cluster r , \mathbf{m}_r is the cluster centre, i.e. the mean. Traditionally the optimal number of clusters is chosen by finding $\max_{k_{opt}} (h_k - h_{k-1})$, known as the L-Curve method. However the problem with this method is that the difference between $(h_k - h_{k-1})$ is not normalised, which may give an incorrect estimate of the optimal number of clusters. The solution to this problem was proposed by the authors in [8] by comparing the clusters to a reference dataset $\mathbf{r}_i = 1, \dots, B$, where B is the total number of reference datasets. The reference dataset is formed by scaling a uniformly distributed random dataset by the range of the principal components of the clusters. Then the reference dataset is clustered and h_{kb} is evaluated, where $b = 1, \dots, B$. The Gap statistic is computed as

$$Gap(k) = \frac{1}{B} \sum_{b=1}^B \log h_{kb} - \log h_k. \quad (5)$$

Where h_{kb} and h_k are the cluster compactness of the reference datasets and EEG data, respectively. The optimum number of clusters is then defined as the smallest k which satisfies

$$Gap(k) \geq Gap(k+1) - s_{k+1}, \quad (6)$$

where

$$s_k = \sqrt{1 + \frac{1}{B}} \sigma_k \quad (7)$$

where σ_k is the standard deviation of $\{\log h_{kb}\}_{b=1, \dots, B}$.

2.3 Reconstruction

Each of the atoms are reconstructed by choosing the data points from the mask that belongs to each cluster. Let M_c denote the mask with one cluster selected. The STFD for the cluster is given by

$$P_c = M_c \cdot P \quad c = 1, \dots, k_{opt} \quad (8)$$

where the space-time-frequency indices have been omitted, (\cdot) is the elementwise multiplication. Next the time series signal is reconstructed by computing the inverse short-time Fourier transform (ISTFT) of $P_c(x, y, t, f)$ defined as

$$A_k(t) = \frac{1}{p} \sum_{l=1}^p \frac{1}{2\pi} \sum_{\omega} \sum_{\tau} w(t - \tau) P_c(x, y, t, \omega) e^{i\omega\tau}. \quad (9)$$

Where $A_k(t)$ $k = 1, \dots, k_{opt}$ is the reconstructed atom, p is the number of electrodes that fall within spatial coordinates of atom k .

2.4 Motion Characterisation

In this section we quantify motion of the sources in order to determine whether there is left or right finger movement. We find the extracted atom's, $A_k(t)$, cross correlation with the raw EEGs over an overlapping window of length L , and with an overlap O . Then the absolute maximum value for each window of cross correlation is used as the location of the atom, given by

$$\rho_k(t) = \max_j (|E\{A_k(t) s_j(t)\}|) \quad (10)$$

and the location (coordinates) is deduced by the index j . For example, if the maximum cross correlation occurred in electrode C_z at $t = 1$ then in C_3 at $t = 2$, the transition would be from coordinates $\rho_k(1) = \{6, 6\}$ to $\rho_k(2) = \{4, 6\}$. Since the atom is disjoint in time space and frequency, there should be only one peak in the cross correlation function for each window. Finally the average direction is given by

$$\mathbf{d}_k = \frac{1}{T} \sum_t \rho_k(t) \quad (11)$$

where \mathbf{d}_k is the direction for atom k , and T is the number of cross correlation windows.

2.5 Classification

We use an SVM as our classifier, due to its generalization and its established empirical performance [9]. The goal of an SVM is to find an optimal separating hyperplane (OSH)

for a given feature set. The OSH is found by solving the following constrained optimisation problem,

$$\begin{aligned} \min_{\mathbf{z}, b, \gamma_{i=1, \dots, l}} & \left(\frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^l \gamma_i \right) \\ \text{s.t. } & q_i (\mathbf{z} \cdot \mathbf{g}_i - b) + \gamma_i \geq 0 \quad i = 1, \dots, l \end{aligned} \quad (12)$$

where $\|\mathbf{z}\|^2 = \mathbf{z}^T \mathbf{z}$ is the squared Euclidean norm and (\cdot) is the dot product. The parameter \mathbf{z} determines the orientation of the separating hyperplane, γ_i is the i -th positive slack parameter, \mathbf{g}_i is a vector containing the features $\mathbf{g}_i = [m_{x_1}(i), m_{y_1}(i), d_{x_{k_1}}(i), d_{y_{k_1}}(i), \dots, m_{x_{k_{opt}}}(i), m_{y_{k_{opt}}}(i), d_{x_{k_{opt}}}(i), d_{y_{k_{opt}}}(i)]^T$, where $m_{x_1}(i)$ and $m_{y_1}(i)$ are the x and y components of the cluster centres, $d_{x_{k_1}}(i)$ and $d_{y_{k_1}}(i)$ are the x and y components of the directional vector. Here, l is the number of training vectors and $q_i \in \{\pm 1\}$ are the output targets. The non negative parameter C is the (misclassification) penalty term, and can be considered as the regularization parameter and is selected by the user. A larger C is equivalent to assigning a higher penalty to the training errors. The parameter C is usually set to a high value to avoid any training error. SVs are the points from the dataset that fall closest to the separating hyperplane. Any vector \mathbf{g}_i that corresponds to a non-zero α_i is a support vector (SV) of the optimal hyperplane. It is desirable to have the number of SVs small to have a more compact and parsimonious classifier. The OSH (generally nonlinear) is then computed by solving (12) using Karush-Kuhn-Tucker conditions [10] as a decision surface of the form

$$f(\mathbf{g}) = \text{sgn} \left(\sum_{i=1}^{L_s} q_i \alpha_i K(\mathbf{g}_i^s, \mathbf{g}) + b \right). \quad (13)$$

In this formula $\text{sgn}(\cdot) \in \{\pm 1\}$, \mathbf{g}_i^s are SVs, $K(\mathbf{g}_i^s, \mathbf{g})$ is the nonlinear kernel function (if $K(\mathbf{g}_i^s, \mathbf{g}) = \mathbf{g}_i^s \cdot \mathbf{g}$ the SVM is linear). A Kernel for a nonlinear SVM projects the samples to a feature space of higher dimension via a nonlinear mapping function. Among nonlinear kernels the radial based function (RBF) defined as $K(\mathbf{g}_i, \mathbf{g}) = \exp(-\|\mathbf{g} - \mathbf{g}_i\|^2 / (2\alpha))$, where the adjustable parameter α governs the variance of the function, is widely used due to having quasi-Gaussian distribution for datasets of large samples.

3. EXPERIMENTS

3.1 Data Collection

The data was provided by King's College Hospital and are available from our website [11]. The EEG was collected using 64 electrodes using Neuroscan. The electrodes were placed using the extended 10-20 system referenced to linked mastoids. During acquisition the electrode impedance was kept below 5k Ω . The signal was sampled at 2kHz and low-pass filtered with a cutoff frequency of 200Hz. An able bodied subject was seated with arms resting on a table and pressed a microswitch approximately every 5 seconds alternating left and right fingers. The data was divided into epochs of 4 seconds, 2 seconds before the movement and 2 seconds after the movement.

3.2 Testing the Algorithm

In our study we tested the features using 100 trials in total; 50 for left finger movement and 50 for right finger movement.

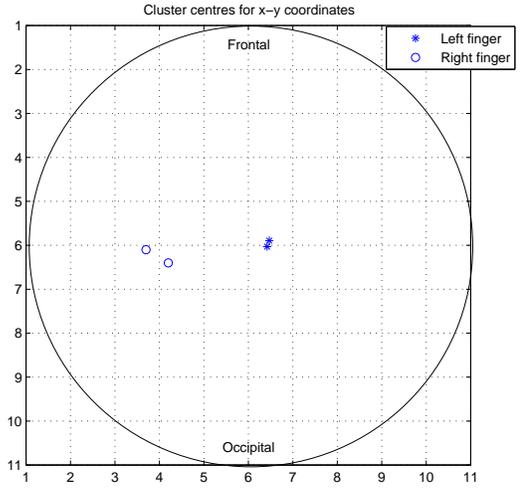


Figure 2: The cluster centres for the extracted atoms for left finger movement.

Table 1: The performance of the classifier based on the average number of correctly classified points. Three kernels are compared in the classification.

Kernel	Average classification rate (%) (s.d.)		
	Overall	Right	Left
Gaus. RBF	75.50 (1.0)	75.16 (1.2)	69.43 (1.5)
Cubic Poly.	65.30 (1.4)	66.15 (1.0)	64.36 (1.0)
Linear	61.01 (1.3)	60.34 (1.4)	56.51 (1.0)

In order to test the overall classification rate we used 4-fold cross-validation (CV) with no overlap, i.e. using 75% of the data for training and 25% for testing. The CV was performed 10 times, with each time the data was chosen at random from our trial pool. The classifier was used with three kernels, linear, RBF and cubic polynomial, for which the error is shown in Table 1. For our dataset the value chosen for the parameter C was 64 and for the case of the RBF kernel the parameter α was set to 0.5. The parameter B was set to 18 reference datasets and the maximum number of clusters, K , was set to 6. We used a Hanning window function for the STFT algorithm. The window length, L , for the motion characterisation algorithm was set to 2000 samples, and the overlap, O , was 1900 samples.

The cluster centres in the spatial domain are shown in Fig. 2. From the figure it can be seen that the cluster centres for the left finger movement occur on the contralateral hemisphere very close to C_2 electrode location, which is located over the motor cortex. For right finger movements the location of the cluster centre is at electrode C_3 , which is associated with right finger movements as explained in [12]. Figure 3 shows the time frequency representation for the atoms of the left finger movement trial. It shows that the clusters are formed from the Alpha band activity. There are two clusters for each trial because of the desynchronisation in the Alpha band during finger movement. At this point the Alpha band power falls below the threshold, μ , and is interpreted as a separate cluster by the kmean clustering algorithm.

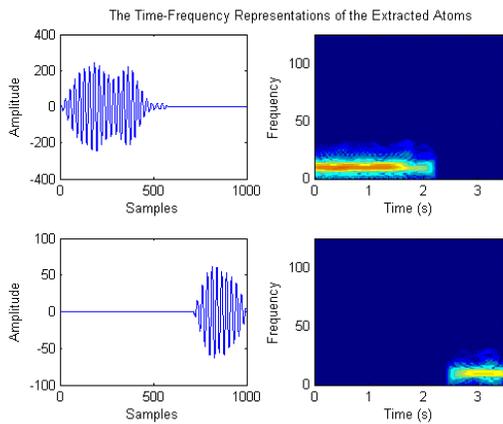


Figure 3: The time-frequency representation of the extracted atoms for a left finger trial.

The average number of support vectors calculated when using the RBF kernel was 35.5% of the training examples. When using the linear kernel the average number of SVs found was 80.2% and for cubic polynomial it was 65.5%. The training error was found by using the training data as test data. The training error was found to be 0.5% (ave.) and the test error was 0.7% (ave.). Since the two errors are close together this gives an indication that overfitting has been avoided.

4. CONCLUSION

We have presented a new method for distinguishing between left and right finger movements from scalp EEGs using the features corresponding to the activity of Alpha rhythms and directionality of the sources. The experiments herein demonstrated that for the test dataset the signals are correctly classified by using the introduced features. Using k-mean clustering followed by the Gap statistic method enables to estimate the number of disjoint factors, representing the brain's active sources, accurately. A higher classification rate is achieved when the RBF kernel is used for the SVM.

REFERENCES

- [1] D. Mcfarland, W. Sarnacki, and J. Wolpaw, "Brain-computer interface (BCI) operation: optimizing information transfer rates," *Biol. Psych.*, vol. 63, pp. 237–251, 2003.
- [2] N. Neumann and N. Birbaumer, "Predictors of successful self control during brain-computer communication," *J Neurol Neurosurg Psychiatry*, vol. 74, pp. 1117–1121, 2003.
- [3] J. Bayliss and D. Ballard, "A virtual reality testbed for brain-computer interface research," *IEEE Trans. on Rehab. Eng.*, vol. 8, no. 2, pp. 188–90, 2000.
- [4] M. Kaminski, M. Ding, W. Truccolo, and S. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol Cybern*, vol. 85, pp. 145–157, 2001.
- [5] J. Ginter-Jr., K. Blinowska, M. Kaminski, and P. Durka, "Phase and amplitude analysis in time-frequency space

- application to voluntary finger movement," *J Neurosc Meth*, vol. 110, pp. 113–124, 2001.

- [6] L. Shoker, S. Sanei, and A. Sumich, "Distinguishing between left and right finger movement from EEG using SVM," *Procs. IEEE Engineering in Medicine and Biology Society 2005, Shanghai*.
- [7] A. Jain and K. Karu, "Learning texture discrimination masks," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 18, no. 2, pp. 195–205, 1996.
- [8] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a dataset via the gap statistic," *Journal of the Royal Statistical Society, B*, vol. 63, pp. 411–423, 2001.
- [9] S. Gunn, "Support vector machines for classification and regression," Univ. of Southampton, Tech. Rep., 1998.
- [10] R. Fletcher, *Practical methods of optimisation, 2nd Ed.* John Wiley, 1987.
- [11] Centre of DSP, Cardiff University website, [Online], Available: <http://www.engin.cf.ac.uk/research/groups/cdsp/index.html>.
- [12] G. Pfurtscheller and C. Neuper, "Synchronization of mu rhythm in the EEG over the cortical hand area in men," *Neurosc Let*, vol. 174, pp. 93–96, 1994.