

# An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain–computer interface

Tao Wang and Bin He<sup>1</sup>

Department of Bioengineering, University of Illinois at Chicago, IL 60607, USA

E-mail: binhe@umn.edu

Received 29 September 2003

Accepted for publication 11 November 2003

Published 20 January 2004

Online at [stacks.iop.org/JNE/1/1](http://stacks.iop.org/JNE/1/1) (DOI: 10.1088/1741-2560/1/1/001)

## Abstract

The recognition of mental states during motor imagery tasks is crucial for EEG-based brain–computer interface research. We have developed a new algorithm by means of frequency decomposition and weighting synthesis strategy for recognizing imagined right- and left-hand movements. A frequency range from 5 to 25 Hz was divided into 20 band bins for each trial, and the corresponding envelopes of filtered EEG signals for each trial were extracted as a measure of instantaneous power at each frequency band. The dimensionality of the feature space was reduced from 200 (corresponding to 2 s) to 3 by down-sampling of envelopes of the feature signals, and subsequently applying principal component analysis. The linear discriminate analysis algorithm was then used to classify the features, due to its generalization capability. Each frequency band bin was weighted by a function determined according to the classification accuracy during the training process. The present classification algorithm was applied to a dataset of nine human subjects, and achieved a success rate of classification of 90% in training and 77% in testing. The present promising results suggest that the present classification algorithm can be used in initiating a general-purpose mental state recognition based on motor imagery tasks.

## 1. Introduction

Among the various theories on the implementation of the brain–computer interface (BCI) (Wolpaw *et al* 2000, 2002, Pfurtscheller and Neuper 2001, Wickelgren 2003), electroencephalographic activities recorded from sensorimotor areas, while a subject is performing a certain motor task, are one of the important schemes worth further investigation.

Oscillation activities at the *mu* and the *beta* rhythm originating in the sensorimotor cortex and associated areas have been found useful in distinguishing mental states as a means to control the directions of a moving cursor on screen (Wolpaw and McFarland 1994). These event-related changes

of EEG activity at specific frequency bands are usually termed event-related desynchronization (ERD) or event-related synchronization (ERS) (Pfurtscheller and Neuper 2001).

It is important to identify robust feature descriptions for BCI research, either in the time or frequency domain as derived from single-trial motor imagery (MI) EEG data. Autoregressive models are commonly used in terms of model parameters without direct relation to specific brain activities (Anderson *et al* 1998, Pfurtscheller *et al* 1998). Time-domain features, such as activity, mobility and complexity, describing the properties of a signal trial have been used (Obermaier *et al* 2001). A direct ERD estimation calculating the energies within pieces of time windows imposed on EEGs at pre-selected frequency bands has been found to be a good feature in classifying right- or left-hand imagined movements (Pfurtscheller *et al* 1997). Another

<sup>1</sup> Present address: Department of Biomedical Engineering, University of Minnesota, MN, USA.

important portion of BCI is the classification method. Learning vector quantization (LVQ), an iterative learning algorithm of labeled code book generation, was modified by weighting the data vector according to the importance of classification (Pfurtscheller *et al* 1997). A so-called local neural classifier—a multilayer perceptron architecture which is made of prototypes associated with different mental tasks—was reported to be preferred to classic models, such as LVQ, and regularized radial basis function networks (Millán *et al* 2002). These nonlinear approaches, however, also have limitations of overfitting, which have to be tackled with regularization methods.

In the present study, we have developed a new feature description based on the rhythmic components of single trial EEGs, and a new weighting strategy determining parameters of the classification algorithm. We have tested the performance of the present algorithm in recognizing the binary MI mental states in nine human subjects.

## 2. Method

### 2.1. Data description

The MI dataset was made available by Dr Allen Osman of the University of Pennsylvania (Osman and Robert 2001) for a data analysis competition during the Neural Information Processing Systems (NIPS 2001) Brain–Computer Interface Workshop (Whistler, Canada, December 2001) (Sajda *et al* 2002).

EEG data were recorded from fifty-nine channels (placed according to the international 10/20 system) with a 100 Hz sampling rate while the subjects were performing synchronized imagined movement tasks. The letter of the electrode name identifies the particular sub-cranial lobe (e.g. F: frontal lobe; C: central lobe). The number or second letter identifies its hemispherical location. The subjects were to synchronize an indicated response with a highly predictable timed cue. A trial was defined as a six-second-long period consisting of several pieces of timing arrangement. Two important timing cues that should be mentioned here are that at 3.75 s of a trial period, a cue (*preparation cue*) of one letter (L or R) appeared on the screen indicating which hand (specifically index finger) movement should be imagined; and at 5.0 s another cue (*execution cue*) appeared, indicating that it was time to make the requested response. The subjects were well trained to consistently respond to the cue signals within 100 ms. The imagined movement is to press a button with the specified index finger. The experiment was designed with both actual movement and imagined movement mixed during each block. Only the imagined left/right trials are processed in the present study.

### 2.2. Spatial filtering

The primary drawback of scalp EEG is the spatial smearing and low signal-to-noise ratio. Spatial filter techniques attempt to address this issue by incorporating spatial information from multi-channel recordings. Since there are correlations among the recording electrodes, spatial filters have been used as a

means of accentuating localized activity and reducing diffused activity. For example, principal component analysis (PCA) was used to remove artifacts (e.g. EOG) (Lagerlund *et al* 1997). This can also be achieved through independent component analysis techniques (Jung *et al* 2000). These methods, however, suffer from the loss of channel coordinate, noncausality and computational inefficiency, which make them unsuitable in an on-line BCI scenario. Another kind of spatial filter re-references the electrode arrangement. McFarland *et al* (1997) have compared the EEG classification results using different spatial filters and concluded that the common average and the Laplacian derivation yield good performance.

The surface Laplacian method (Hjorth 1975, Perrin *et al* 1987, Nunez *et al* 1994, Babiloni *et al* 1996, He 1999, He *et al* 2001), which derives the second spatial derivative of the instantaneous spatial potential distribution, could serve for such a high-pass spatial filtering purpose. Assuming that the distances from a given electrode to its four directional neighboring electrodes are approximately equal, the surface Laplacian can be approximated by subtracting the average value of the neighboring channels from the channel of interest (Hjorth 1975), i.e.

$$V_j^{\text{Lap}} = V_j - \frac{1}{4} \sum_{k \in S_j} V_k, \quad (1)$$

where  $V_j$  is the scalp potential EEG of the  $j$ th channel, and  $S_j$  is an index set of the four neighboring channels. The above-defined surface Laplacian is usually referred to as the local estimate of the surface Laplacian and was used in the present study due to its computational simplicity and efficiency as compared to other surface Laplacian estimation methods.

### 2.3. Frequency decomposition

Fifth order IIR Butterworth band-pass filters were used for frequency decomposition. The filter coefficients were derived such that the frequency band from 5 to 25 Hz is evenly divided into 20 band bins with approximately 2 Hz bandwidth and 50% band overlapping. These frequency bands were chosen because *mu* and *beta* rhythm components were reported to be the significant frequency characteristics reflecting MI tasks (Pfurtscheller and Neuper 2001). The optimal frequency bands, however, vary from subject to subject. A training procedure, therefore, is required to grade the contribution of each frequency band to the classification performance. A similar idea was implemented in an EEG classification algorithm—known as distinction sensitive learning vector quantization (DSLQ)—which used a weighted distance function and adjusted the influence of different input features through a supervised learning algorithm (Pregenzer *et al* 1996).

The ERD/ERS features used in BCI can be derived by squaring temporally filtered EEGs at the subject-specified optimal frequency bands and then averaged in time to smooth the waveforms. Transient variation, thus, might be overridden. The power spectrum estimated by parametric models or a modified periodogram, on the other hand, cannot localize such frequency features in time.

Through the narrow band-pass filtering, the EEG signals were decomposed into groups of different frequency components. This procedure results in a time signal containing a rapid oscillation modulated by amplitude variation, which can be approximately expressed as

$$x(t) = a(t) \cos(2\pi f_0 t + \varphi(t)), \quad (2)$$

where  $a(t)$  is a low-frequency signal that contains event-related activities that are time-locked to the cue and associated with the MI intention,  $f_0$  is the frequency of the modulating signal and  $\varphi(t)$  is the initial phase. The Hilbert transform of (2) can be obtained as (Papoulis 1977)

$$x_h(t) = a(t) \sin(2\pi f_0 t + \varphi(t)), \quad (3)$$

and the corresponding analytical signal is

$$z(t) = x(t) + jx_h(t) = a(t) \exp(2\pi f_0 t + \varphi(t)). \quad (4)$$

Therefore, the envelope can be estimated as

$$a(t) = |z(t)| = \sqrt{x^2(t) + x_h^2(t)}. \quad (5)$$

The ERD/ERS phenomenon can be observed through the grand average of the envelopes over trials. For a single trial and a narrow band EEG signal, its envelope contains the relevant information for instantaneous power estimation (figure 1).

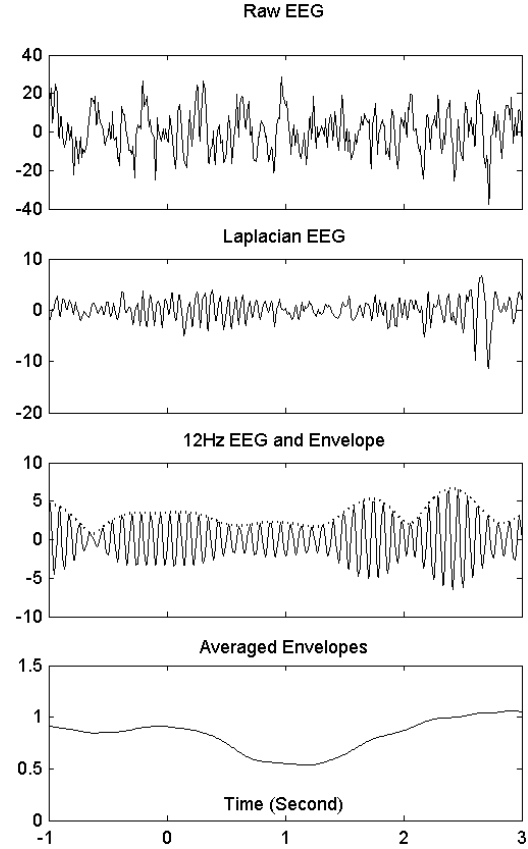
Usually, the envelopes are in a very low frequency band. Down-sampling will substantially reduce feature dimensions while retaining the original information. This would also substantially reduce the computation time required in the following steps. The sampling rate, therefore, was reduced by a factor of 10 because the frequency band for most envelopes is below 5 Hz.

The dynamic region of the EEG amplitude would be dramatically altered due to its non-stationary nature. In particular, the spatial noise would be amplified by the surface Laplacian operation. Thus, an amplitude normalization operation was performed to address this concern, by making the L2 norm of a feature vector (the envelop signal) a constant.

#### 2.4. Linear classification

A linear classifier was used in the present study, which maximizes the ratio of inter-class variance to the intra-class variance in any particular datum, thereby guaranteeing maximal separability. Linear discriminate functions, however, cannot reliably be estimated in the high dimensional feature space and with the small sample size. PCA is an optimal method, in the sense of least mean squared reconstruction error, to reduce the dimensionality of the feature space since it reduces a large number of variables to fewer components that are orthogonal to one another and account for the maximal amount of variance in the original data. Lugger *et al* (1998) investigated the classification scheme based on dimensionality reduction by means of PCA for two sets of features, and concluded that the principal components accounting for high variance are not the optimal candidates for classification tasks. This, however, is only true for non-Gaussian cases.

A two-second period of the enveloping EEG starting from the preparation cue was selected (corresponding to 20 points after down-sampling) as the underlying motor



**Figure 1.** The evolution of a piece of raw EEG (first row) through a series of EEG processing steps. Time 0 indicates the preparation cue onset moment. Values from 0 to 2 s were used for classification. The second row shows Laplacian filtered EEG signals. Rhythmic power variations along the time course are delineated by its envelopes (third row), which constitutes an original feature description. By averaging these envelopes of one mental state over all trials, clear ERD/ERS patterns can be observed (fourth row).

imagery duration. By using PCA on the two-dimensional matrix composed of period  $\times$  trial in the training set, the feature vectors projecting onto the three largest principal components were retained.

Let  $v$  be a feature vector obtained from previous PCA procedures of one trial and  $h_{i,j}(v)$ ,  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, M$  be the *linear discriminant function* for the  $j$ th channel and the  $i$ th band. Thus the decision rule is that if  $h_{i,j}(v) < 0$ , then  $v$  belongs to the class of ‘left’ trials or if  $h_{i,j}(v) \geq 0$ , then  $v$  belongs to the ‘right’ trial.

During the training stage, approximately the same number of left and right imagined movements were provided to each linear classifier  $h_{i,j}(v)$ , and the correct classification accuracy, denoted by  $a_{i,j}$ , was obtained. We hypothesized that these accuracy values reflect the adaptation of every frequency band and channel, and defined a normalized *frequency weight* for each channel as

$$w_{i,j} = \begin{cases} (2a_{i,j} - 1)^m, & a_{i,j} > 0.5 \\ 0, & a_{i,j} \leq 0.5, \end{cases} \quad (6)$$

where  $m$  is a control parameter used to further emphasize those bands with larger accuracy values (here  $m = 2$ ).

**Table 1.** Classification accuracy (%) for each subject using two pairs of channels.

Subjects	1	2	3	4	5	6	7	8	9	Mean $\pm$ Std
Training	85.3	93.0	83.0	83.3	82.4	87.5	88.6	80.7	88.6	85.8 $\pm$ 3.9
Testing	75.6	85.6	70	68.9	70.0	80.0	82.2	67.8	78.3	75.4 $\pm$ 6.5

**Table 2.** Classification accuracy (%) for each subject using four pairs of channels.

Subjects	1	2	3	4	5	6	7	8	9	Mean $\pm$ Std
Training	88.2	94.8	89.0	89.0	85.6	89.0	95.0	86.6	94.8	90.2 $\pm$ 3.7
Testing	80	90.6	70.6	62.8	73.9	79.4	92.2	69.4	73.9	77.0 $\pm$ 9.7

According to the frequency weight, a synthetic decision for one channel over all frequency bands was made by the following *channel discriminant function*:

$$g_j(v) = \sum_{i=1}^N w_{i,j} \operatorname{sgn}(h_{i,j}(v)). \quad (7)$$

The correct classification accuracy for the  $j$ th channel of all frequency bands was obtained using this method. So the decision for  $v$  was made according to the sign of  $g_j(v)$ , and its absolute value measures the likelihood of this decision. The logical way of synthesizing all channels is summarizing the  $g_j(v)$  together as

$$q(v) = \sum_{j=1}^M g_j(v). \quad (8)$$

Therefore, the final classification decision was made by synthesizing all frequency bins and channels according to their classification contribution as rated by training.

### 3. Results

#### 3.1. Classification accuracy

There were two channel combinations—two pairs of channels (C1/2, C3/4) and four pairs of channels (plus FC1/2 and FC3/4)—tested in the present study. The data of all 180 trials per subject were split into training and testing sets by 10-fold cross validation. The number of two groups (left and right) of EEG patterns was set to be even in each fold so as to obtain unbiased training discrimination weights. Tables 1 and 2 present the accuracies (i.e. successful classification rate) for the two combinations respectively. The mean testing accuracies of nine subjects for two situations are virtually equal while the second combination shows larger variation among subjects. Seven out of nine subjects yield better or virtually the same scores for more channels involved while subjects 4 and 9 deteriorate around 5% on average. For the training set, the accuracies are more than 10% higher than that of the testing set.

This can be explained by examining the generalization performance of an individual channel. Figure 2 shows the classification accuracy of individual channels (32 in total) for training (dot line) and testing (solid line) for two subjects, while figure 3 shows the weight distribution of four pairs of channels for two subjects. Usually C3 and C4 possess not

only higher training accuracy but also strong generalization capability for most subjects. For other channels, however, such capability varies across subjects. So the results should be improved by using more channels because they take into account more information for some subjects. We might, however, also take the risk of decreased performance when bad generalization channels are taken.

Take a typical example of subject 7. As figure 2(a) shows, channels in the bilateral central sulcus area (C1–4) as well as the bilateral frontal central area (FC1–4) all possess high generalization capability and thus four pairs of channels increase the success rate by 10%. In contrast, for subject 4, the strong generalization channels are in the parietal central (channels CPx,  $x$  denotes the available suffix of the channel names) area rather than FCx, which resulted in a decrease of around 6% (figure 2(b)).

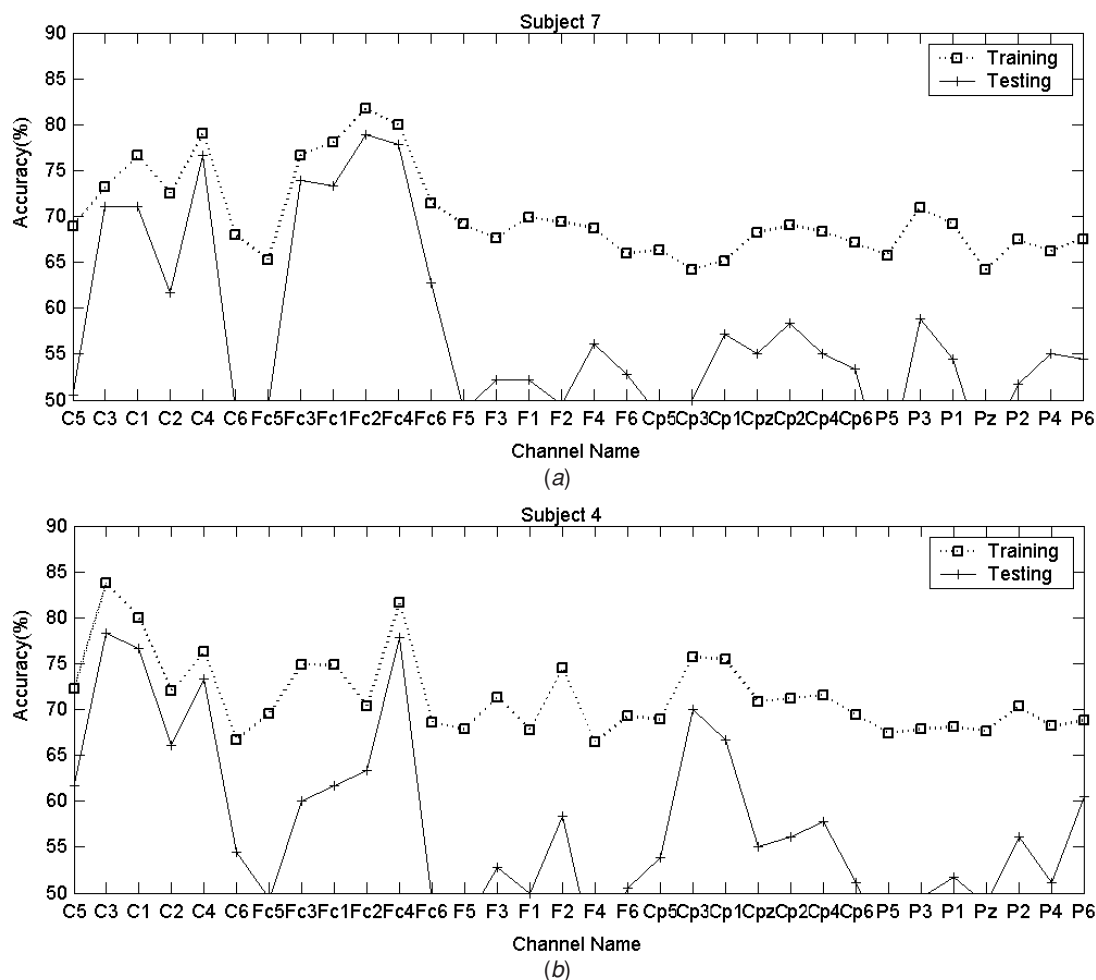
On the other hand, from figure 2 we note that it is not possible to select channel combinations purely based on sorting channels yielding higher accuracies at the training stage. In practice, based on the current framework, subject specific channel combination is viable by the trial and error method. For example, if we try to replace the FCx channels with the other four channels (e.g. F4, F6, CP5 and CP1) in subject 4, the score could be increased from 62.8 to 72.2%.

#### 3.2. Frequency analysis

During the training stage, there are two frequency ranges at the *mu* and the *beta* band that were picked up. An instance of frequency weight distribution in figure 3 shows the significant band bins for classification. For those channels with bad performance such distribution patterns deteriorated substantially. Note that the relevant frequency bands are not limited to those with obvious ERD/ERS, but also those bands where the difference between ERD/ERS of left and right MI is larger (see figure 4). Figure 4 shows a clearly contralateral ERD phenomenon at this significant frequency band, being consistent with previous findings (Babiloni *et al* 2001). There are also frequency bands that have a strong ERD phenomenon but provide little discrepancy between left and right MI.

### 4. Discussion

Single trial EEGs contain strong artifacts from various sources, which can be decomposed into different bands by frequency



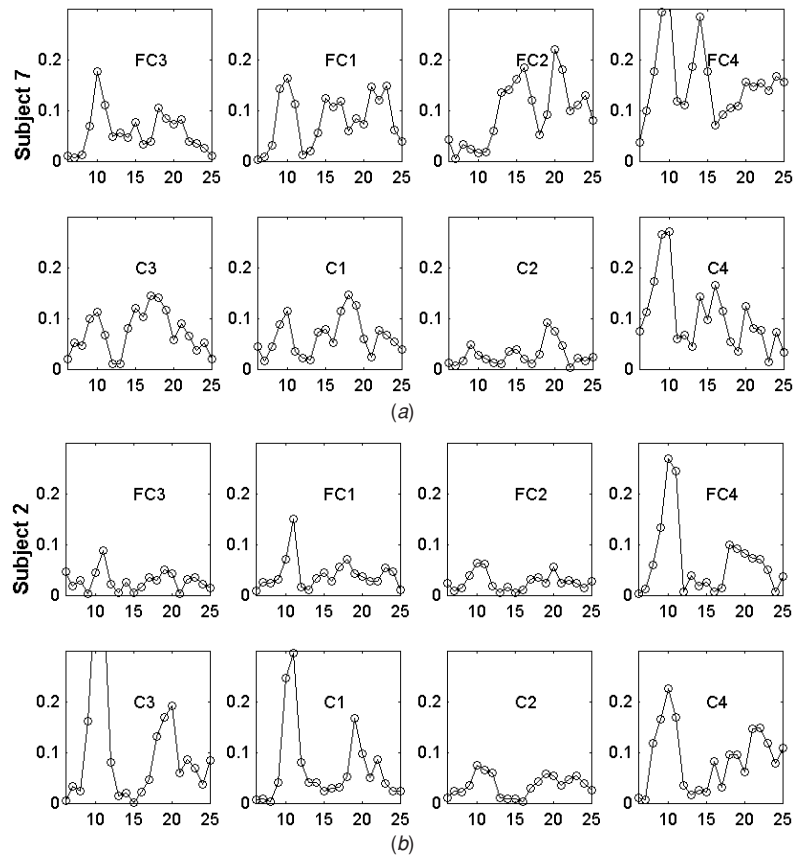
**Figure 2.** Classification accuracy of individual channels (32 in total) for training (dot line) and testing (solid line) for subject 7 (a) and subject 4 (b). Note that classification accuracy for training is always better than that of testing. Also note that substantial variation exists from channel to channel for testing, which suggests the different generalization capability of the different channels.

decomposition. Those frequency bins that possess high signal-to-noise ratio can be picked up by assessing their contributions to classification. On the other hand, envelopes at each band are essentially a morphological pattern in the time domain and time-locked to the MI events. A BCI could conceivably use features in both the time domain and frequency domain and may thereby improve performance. However, for a single frequency bin, the classification success rate is still limited (less than 65% for most cases). Incorporating multiple frequency bins by using the present synthesis strategy further improves the classification success rate (usually by 5%). Furthermore, improvement may be made by synthesizing contributions from selected multi-channel signals. After applying the surface Laplacian filter, the signals from neighboring channels may be assumed not to be strongly correlated. For example, the correlation coefficients between selected channels (C3, C4 for instance) and their neighboring channel were significantly reduced from 0.93 to 0.25 in average for the nine subjects studied. When the cross-channel correlation is low, the multichannel classifiers would synthesize different information. However, as we have seen

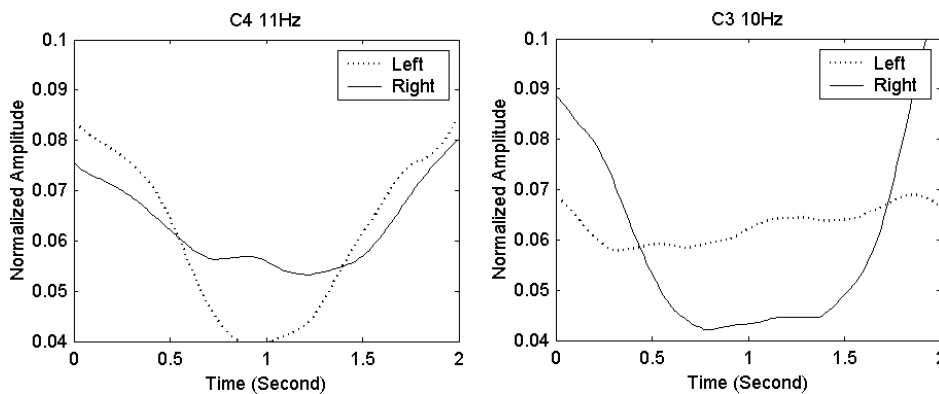
from figure 2, the generalization capacities for each channel vary, thus the testing performance may not improve when more channels are included.

In addition to linear discrimination, a quadratic classifier was also tried in this study. The classification accuracy for the training set improved but decreased for the testing set. Due to the small training set where large deviations are possible and overfitting might occur, a linear function that explains most of the data seems to be preferable to the quadratic function. The upper bound accuracy of this linear classification procedure is 85.5% for two pairs of channels, and 90.1% for four pairs of channels, when all the data were used as the training set.

It should be kept in mind that for the imagined movement, it is difficult to know whether or not the desired tasks are fulfilled as the training process intended. There might be three or even more mental states that might occur during the processed time windows. For some trials it is possible that the subjects did ‘nothing’ or at least did not act strictly according to the presented cues. Therefore, it would have been more accurate to reflect the experimental data and achieve a higher classification rate, if such a third ‘state’ had been introduced.



**Figure 3.** Distributions of frequency weights of four pairs of channels for (a) subject 7 and (b) subject 2. The horizontal axis shows the frequency bins in Hz. Note the distribution of frequency weights varies from channel to channel. In general, *mu* and *beta* rhythms show positive contributions to the classification, being consistent over most channels.



**Figure 4.** ERD waveforms obtained by averaging envelopes over trials in a training set for the left and right motor imagery at the most pronounced frequency bins for classification. Note an obvious ERD phenomenon occurs at the contralateral hemisphere (C3 versus right MI and C4 versus left MI).

A practical system should be able to distinguish the target movement and a few deviate trials. Penny *et al* (2000) proposed an idea of incorporating an uncertainty parameter which allows us to estimate the uncertainty associated with each subsequent classification, i.e., some trials could be rejected by the classifier to improve the classification accuracy. Most results reported in BCI literature for classifying overt or imagined movement of limbs (including hands and feet)

ranged from less than 65 to more than 95% for individuals (see e.g. Millán *et al* (2002), Babiloni *et al* (2001), Ramoser *et al* (2000), Pfurtscheller *et al* (1997)). The score depends heavily on the subjects being studied, overt or imagined movements, and the number of tasks (usually two or three), so it is rather difficult to use the comparison of accuracy score to provide full insight into the different method's capabilities. It is worth noting, however, that no online feedback is available in

the present study which prevents further improvement of the extraction and translation algorithm. Since adaptation is an important factor in BCI research (Wolpaw *et al* 2002), online training would probably provide an optimal means of adjusting the parameters being used in a BCI. Further, no attempt is made in the present study to reject any trial, but rather using all trials provided in the dataset. With online training and rejecting of bad trials based on certain criteria (e.g. Penny *et al* 2000), the classification rate would be further improved.

Muscle activity and eye movement can deteriorate or contribute to the electrical activity recorded from the scalp. At frontal, temporal, and occipital locations particularly, EMG and/or EOG can exceed EEG, even in the characteristic EEG frequency bands. While data are not available in the present study, it would be desirable if continuous EEG signals could be monitored and bad trials be rejected, to further improve the system reliability and performance.

In summary, we have developed a new algorithm for the classification of motor imagery tasks using a frequency decomposition and weighted-synthesis method. The present approach does not require a subject-dependent setting in the parameter selection in the time domain, frequency domain and electrode montage. The present promising results in nine human subjects suggest that the present approach merits further investigation, and may become an alternative to initiating a general-purpose mental state recognition system based on motor imagery tasks.

## Acknowledgments

The authors are grateful to Dr Allen Osman of the University of Pennsylvania for making his data available, and Anirudh Vallabhaneni for proofreading the manuscript. This work was supported in part by NSF BES-0218736, NSF CAREER Award BES-9875344 and NIH RO1EB00178.

## References

- Anderson C W, Stolz E A and Shamsunder S 1998 Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks *IEEE Trans. Biomed. Eng.* **45** 277–86
- Babiloni F, Babiloni C, Carducci F, Fattorini L, Onorati P and Urbano A 1996 Spline Laplacian estimate of EEG potentials over a realistic magnetic resonance-constructed scalp surface model *Electroenceph. Clin. Neurophysiol.* **98** 363–73
- Babiloni F, Cincotti F, Bianchi L, Pirri G, Millán J d R, Mouriño J, Salinari S and Marciani M G 2001 Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers *Med. Eng. Phys.* **23** 323–8
- He B 1999 Brain electric source imaging: scalp Laplacian mapping and cortical imaging *Crit. Rev. Biomed. Eng.* **27** 149–88
- He B, Lian J and Li G 2001 High-resolution EEG: a new realistic geometry spline Laplacian estimation technique *Clin. Neurophysiol.* **112** 845–52
- Hjorth B 1975 An on-line transformation of EEG scalp potentials into orthogonal source derivations *Electroenceph. Clin. Neurophysiol.* **9** 526–30
- Jung T P, Makeig S, Humphries C, Lee T W, McKeown M J, Iragui V and Sejnowski T J 2000 Removing electroencephalographic artifacts by blind source separation *Psychophysiology* **37** 163–78
- Lagerlund T D, Sharbrough F W and Busacker N E 1997 Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition *J. Clin. Neurophysiol.* **14** 73–82
- Lugger K, Flotzinger D, Schlögl A, Pregenzer M and Pfurtscheller G 1990 Feature extraction for on-line EEG classification using principal components and linear discriminants *Med. Biol. Eng. Comput.* **36** 309–14
- McFarland D J, McCane L M, David S V and Wolpaw J R 1997 Spatial filter selection for EEG-based communication *Electroenceph. Clin. Neurophysiol.* **103** 386–94
- Millán J R, Mouriño J, Franzé M, Cincotti F, Varsta M, Heikkinen J and Babiloni F 2002 A local neural classifier for the recognition of EEG patterns associated to mental tasks *IEEE Trans. Neural Networks* **13** 678–86
- Nunez P, Silibertein R B, Cdush P J, Wijesinghe R S, Westdrop A F and Srinivasan R 1994 A theoretical and experimental study of high resolution EEG based on surface Laplacian and cortical imaging *Electroenceph. Clin. Neurophysiol.* **90** 40–57
- Obermaier B, Guger C, Neuper C and Pfurtscheller G 2001 Hidden Markov models for online classification of single trial EEG data *Pattern Recognit. Lett.* **22** 1299–309
- Osman A and Robert A 2001 Time-course of cortical activation during overt and imagined movements *Cognitive Neuroscience Annual Meeting (New York, 2001 March)*
- Papoulis A 1977 *Signal Analysis* (New York: McGraw-Hill)
- Penny W D, Roberts S J, Curran E A and Stokes M J 2000 EEG-based communication: a pattern recognition approach *IEEE Trans. Rehabil. Eng.* **8** 214–5
- Perrin F, Bertrand O and Pernier J 1987 Scalp current density mapping: value and estimation from potential data *IEEE Trans. Biomed. Eng.* **34** 283–8
- Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer communication *Proc. IEEE* **89** 1123–34
- Pfurtscheller G, Neuper C, Flotzinger D and Pregenzer M 1997 EEG-based discrimination between imagination of right and left hand movement *Electroenceph. Clin. Neurophysiol.* **103** 642–51
- Pfurtscheller G, Neuper C, Schlögl A and Lugger K 1998 Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters *IEEE Trans. Rehabil. Eng.* **6** 316–25
- Pregenzer M, Pfurtscheller G and Flotzinger D 1996 Automated feature selection with a distinction sensitive learning vector quantizer *Neurocomputing* **11** 19–29
- Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6
- Sajda P, Gerson A, Müller K R, Blankertz B and Parra L 2003 A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces *IEEE Trans. Neural Systems Rehabil. Eng.* **11** 184–5
- Wickelgren I 2003 Tapping the mind *Science* **299** 496–9
- Wolpaw J R and McFarland D J 1994 Multichannel EEG-based brain-computer communication *Electroencephalogr. Clin. Neurophysiol.* **90** 444–9
- Wolpaw J R, Birbaumer N, Heetderks W J, McFarland D J, Peckham P H, Schalk G, Donchin E, Quatrano L A, Robinson C J and Vaughan T M 2000 Brain-computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164–73
- Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91