A Time-Frequency Approach to Feature Extraction for a Brain-Computer Interface with a Comparative Analysis of Performance Measures

Damien Coyle

Intelligent Systems Engineering Laboratory, School of Computing and Intelligent Systems, Faculty of Engineering,
University of Ulster, Magee Campus, Derry BT48 7JL, UK
Email: dh.coyle@ulster.ac.uk

Girijesh Prasad

Intelligent Systems Engineering Laboratory, School of Computing and Intelligent Systems, Faculty of Engineering,
University of Ulster, Magee Campus, Derry BT48 7JL, UK
Email: g.prasad@ulster.ac.uk

T. M. McGinnity

Intelligent Systems Engineering Laboratory, School of Computing and Intelligent Systems, Faculty of Engineering,
University of Ulster, Magee Campus, Derry BT48 7JL, UK
Email: tm.mcginnity@ulster.ac.uk

The paper presents an investigation into a time-frequency (TF) method for extracting features from the electroencephalogram (EEG) recorded from subjects performing imagination of left- and right-hand movements. The feature extraction procedure (FEP) extracts frequency domain information to form features whilst time-frequency resolution is attained by localising the fast Fourier transformations (FFTs) of the signals to specific windows localised in time. All features are extracted at the rate of the signal sampling interval from a main feature extraction (FE) window through which all data passes. Subject-specific frequency bands are selected for optimal feature extraction and intraclass variations are reduced by smoothing the spectra for each signal by an interpolation (IP) process. The TF features are classified using linear discriminant analysis (LDA). The FE window has potential advantages for the FEP to be applied in an online brain-computer interface (BCI). The approach achieves good performance when quantified by classification accuracy (CA) rate, information transfer (IT) rate, and mutual information (MI). The information that these performance measures provide about a BCI system is analysed and the importance of this is demonstrated through the results.

Keywords and phrases: brain-computer interface, neuromuscular disorders, electroencephalogram, time-frequency methods, linear classification.

1. INTRODUCTION

Nearly two million people in the United States [1] are affected by neuromuscular disorders. A conservative estimate of the overall prevalence is that 1 in 3500 of the world's population may be expected to have a disabling inherited neuromuscular disorder presenting in childhood or in later life [2]. In many cases those affected may have no control over muscles that would normally be used for communication. BCI technology is a developing technology but has the potential to contribute to the improvement of living standards for these people by offering an alternative communication channel which does not depend on the peripheral nerves or muscles [3]. A BCI replaces the use of nerves and muscles and the movements they produce with electrophysiological signals in conjunction with the hardware and software that translate those signals into actions [1].

A BCI involves extracting information from the highly complex EEG. This is usually achieved by extracting features from EEG signals recorded from subjects performing specific mental tasks. A class of features for each mental task is usually obtained from signals, prerecorded whilst a subject performs a number of repetitions of each mental task. Subsequently a classifier is trained to learn which features belong to which class. This ultimately leads to the development of a BCI system that can determine which mental tasks are related to specific EEG signals [4] and associate those EEG signals with the user's intended communication.

This work demonstrates the use of the short time Fourier transform (STFT) to extract reliable features from EEG signals altered by imagined right/left-hand movements. EEG data was recorded from two recording sites on the scalp positioned at C3 and C4 [5] over the motor cortex. The STFT is used to calculate frequency spectra from a window (i.e., STFT-window) which slides along the data contained within another window (i.e., the feature extraction (FE) window). All EEG data recorded from each recording site is passed through the FE window. The spectra are smoothed using an interpolation (IP) process. Features are obtained from each interpolated spectrum by calculating the norm of the power in predetermined subject-specific frequency bands. Linear discriminant analysis (LDA) is used for classification and system performance is quantified based on three performance measures. The measurement of BCI performance is very important for comparing different systems and measuring improvements in systems. There are a number of techniques used to quantify the effectiveness and performance of a BCI system. These include measuring the classification accuracy (CA) and/or measuring the information transfer (IT) rate. The latter performance quantifier takes into consideration the CA and the time (CT) required to perform classification of each mental task. A third and relatively new quantifier of performance for a BCI system is to quantify the mutual information (MI) which is a measure of the average amount of information a classifier output contains about the input signal [6, 7]. A critical analysis of the performance measures, illustrating the advantages of utilising each one for evaluating a BCI system, is provided.

The performance of the system is dependent upon choices of parameter combinations. It is shown that the width of the main FE window, the number of STFT windows, the width and length of the STFT windows, and the amount of overlap between consecutive STFT-windows all have significant affects on the performance of the system. An interpolation process for smoothing the frequency spectra improves the features and helps increase CA rates. The importance of each parameter is analysed. The results demonstrate that, to obtain the best performance, the parameter combinations have to be optimised individually for each subject. However, a number of parameters converge to similar values, therefore there may exist a particular parameter combination that would generalise well to all subjects and thus potentially simplify the application of the system to each individual subject. Details on these aspects of the system, along with a comparison to other BCI systems, are discussed.

The paper is organised in 11 sections. Section 2 describes the data acquisition procedure. Section 3 introduces the STFT and the FEP and Section 4 provides an analysis of the EEG used in this work. Sections 5 and 6 describe the FEP and the classification procedures, respectively. Section 7 describes briefly three methods for quantifying the performance of a BCI system. Section 8 outlines the system optimisation procedure. Sections 9 and 10 document and discuss the results. Section 11 concludes the paper.

2. DATA ACQUISITION

The EEG data used to demonstrate this approach was recorded by the Graz BCI research group (see acknowledgement) [8, 9, 10, 11]. The Graz group has developed a BCI which uses μ (8–12 Hz) and central β (18–25 Hz) EEG rhythms recorded over the motor cortex. Several factors have suggested that μ and/or β rhythms may be good signal features for EEG-based communication. These signals are associated with those cortical areas most directly connected to the brain's normal motor output channels [1]. The data was recorded from 3 subjects (S1, S2, and S3) over two sessions, in a timed experimental recording procedure. Each trial was 8 s length. The first 2 s was quiet, at $t = 2$ s an acoustic stimulus signifies the beginning of a trial, and a cross "+" was displayed for 1 s, then at $t = 3$ s, an arrow (left or right) was displayed as cue. At the same time the subject was asked to move a bar in the direction of the cue by imagining moving the left or right hand. The feedback (bar movement) can help the user learn to control their EEG better for specific tasks. For subject S1 a total of 280 trials were recorded (140 trials of each type of movement imagery). For the subject S2 there were 320 trials (160 trials of each type of movement imagery). The recording was made using a g.tec amplifier (http://www.gtec.at/) and Ag/AgCl electrodes. All signals were sampled at 128 Hz and filtered between 0.5 and 30 Hz. Two bipolar EEG channels were measured using two electrodes positioned 2.5 cm posterior ("−") and anterior ("+") to position C3 and C4 according to the international standard (10/20 system) electrode positioning nomenclature. In bipolar recording the recorded voltage is the voltage difference between the anterior and posterior electrode at each recording site. A detailed description of similar experimental setups for recording these EEG signals is available [6, 8, 9, 10, 11, 12].

3. THE FE WINDOW AND THE STFT WINDOW

In this investigation there are two windows utilised—the FE window and the STFT window. EEG signals (or data) are fed through the FE window and within the FE window the frequency components of the EEG signal are obtained using a fast Fourier transform (FFT). Within the FE window a temporal resolution is attained by sliding the STFT window along the data sequence with a certain overlap. This windowed signal processing technique is often referred to as the Gabor transform after Gabor (1946). STFT analysis of a nonstationary signal assumes stationarity over the selected signal segment (the STFT window). The inherent assumption of stationarity over the STFT window can lead to smearing in the frequency domain and decreased frequency resolution when analysing EEG signals with fast changing spectral content [13]. The temporal resolution can be made as high as possible by sliding the STFT window along the FE window with a large overlap. This maximises the potential for identifying short events that occur within the FE window [14].
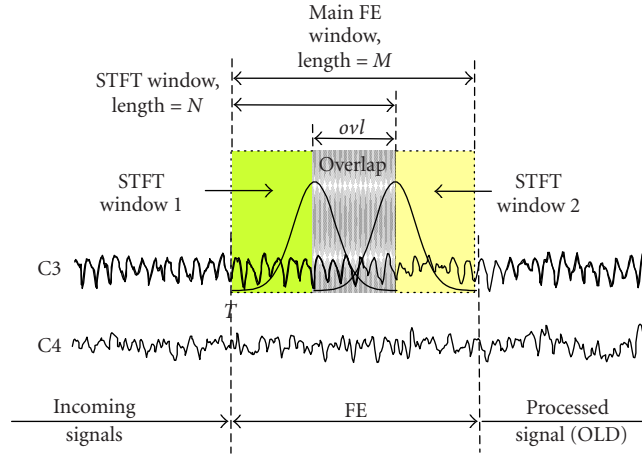
FIGURE 1: Illustration of FE window and STFT window in the FEP.

To localise the Fourier transform of the signal at time instant τ which falls within the main FE window, the STFT-window function is peaked around τ and falls off, thus emphasising the signal in the vicinity of time τ and suppressing it for distant times [15]. There are a number of windows which can be used for achieving these characteristics. Gabor proposed the use of a Gaussian window formulated as follows:

$$w(t) = e^{-(1/2)(\alpha(t-N/2)/(N/2))^2}, \quad (1)$$

where $0 \leq t < N$ and α is the reciprocal of the standard deviation. The width of the window is inversely related to the value of α; a larger α produces a narrower window. The window has the length N. These constant parameters denote the length of the window and the degree of localisation in the time domain, respectively [15]. The tuning of these parameters is very important for the extraction of features used in this approach and this is made apparent in the results section.

The ordinary Fourier transform (FT) is based on comparing the signal with complex sinusoids that extend through the whole time domain; its main disadvantage is the lack of information about the time evolution of the frequencies. In this case, if an alteration occurs at some time boundary, the whole Fourier spectrum will be affected [15]. The FT requires stationarity of the signal which is a disadvantage in EEG analysis, the EEG signal being highly nonstationary. The STFT helps to overcome many of these disadvantages and is formulated as follows:

$$Y_k(f,\tau) = \sum_{i=\tau-N/2}^{\tau+N/2} w^*(t-\tau)y_k(t)e\left(-j\frac{2\pi}{N}ft\right), \quad (2)$$

where $f = 0,1,\ldots,N_f - 1$. N_f is the number of frequency points or Fourier transforms. $Y_k(f,\tau)$ contains the frequency spectrum for each STFT window centred at τ. y_k is the input EEG signal (i.e., either C3 ($k = 1$) or C4 ($k = 2$)) contained within the main FE window. The number of STFT windows used to analyse the data contained in the FE window depends on the length of the FE window M, the STFT window length N, and the amount of overlap, ovl, between adjacent STFT windows. M must always be larger than N. Y_k is a matrix with N_f rows and $E = (M - ovl)/(N - ovl)$ columns (i.e., the rows contain the power of the signal for each harmonic and E is the number of STFT windows that are produced within the FE window).

This analysis was carried out offline although, to approximate the online capabilities, all features are extracted within the FE window so that features can be extracted at the rate of the sampling interval as data passes through the window. As each new signal sample enters the FE window, the oldest sample is removed and the STFT window slides along the signal within the FE window (this process is repeated as each new sample enters the FE window). A frequency spectrum is calculated for each STFT window centred at τ. An illustration of the FE window and STFT window is shown in Figure 1. This illustration shows two STFT-windows contained within the FE window for each signal (C3 or C4).

4. SPECTRAL ANALYSIS AND ERD/ERS

The spectra of signals recorded from recording sites C3 and C4 when subjects perform imagination of hand movements usually show an increase and decrease in the intensity of frequencies in the μ (8–12) and central β (18–25) ranges, depending on the recording location and the imagined hand movement (left or right). When certain cortical areas, such as the sensorimotor area, become activated during the course of information processing, amplitude attenuation occurs in the oscillations of the μ and central β rhythms. This is known as an event-related desynchronisation (ERD). An amplitude enhancement or event-related synchronisation (ERS) can be observed in cortical areas that are not specifically engaged in a given mode of activity at a certain moment of time [9, 11]. The location and frequency ranges of ERS/ERD are subject-specific.
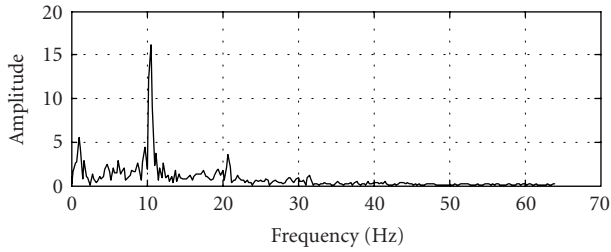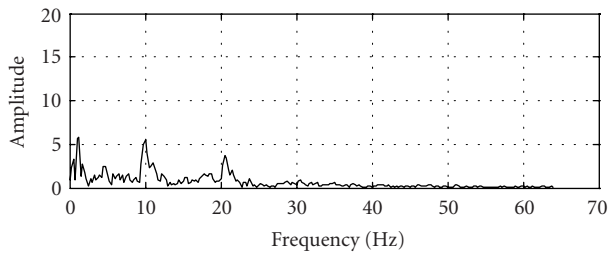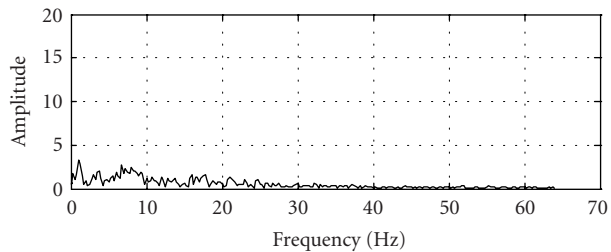
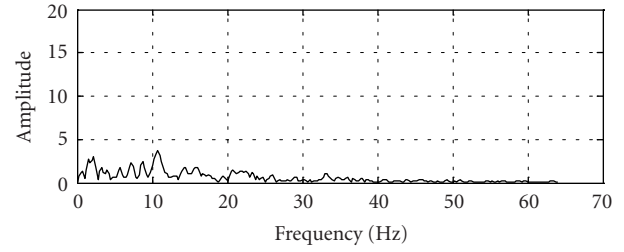Raw spectrum
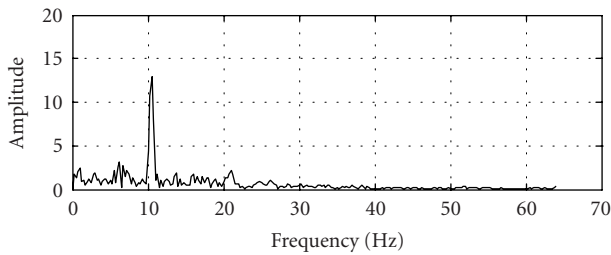
Raw spectrum

Figure 2: C3 Left (windows 1 & 2).

Raw spectrum

Raw spectrum

Figure 4: C3 Right (windows 1 & 2).

Raw spectrum

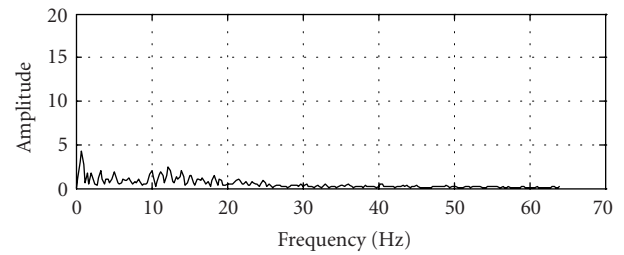Raw spectrum

Figure 3: C4 Left (windows 1 & 2).
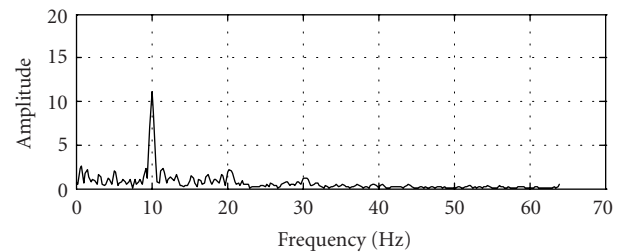
Raw spectrum

Raw spectrum

Figure 5: C4 Right (windows 1 & 2).

Figures 2, 3, 4, and 5 show a typical set of frequency spectra. Figures 2 and 3 are obtained from calculating the STFT from EEG signals recorded during imagination of left-hand movement. Figures 4 and 5 were obtained from signals recorded during imagination of right-hand movement. For this analysis only two windows were used for each signal. The top graph in each figure is the spectrum of the first window and the bottom is the spectrum calculated for the second window. The dominant frequency components can be observed from each graph.
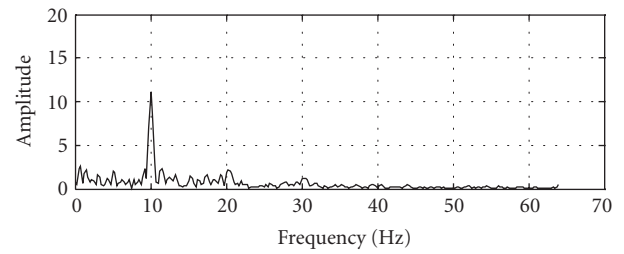
Both spectra in Figure 5 (C4 recording right signal) show strong evidence of μ (8–12 Hz) rhythm. This is not observable from Figure 4 (C3 recording right signal) which suggests that there is an ERD of the μ rhythm on the contra lateral side (opposite side to imagined hand movement). ERD can be interpreted as an electrophysiological correlate of activated cortical areas involved in processing of sensory or cognitive information or production of motor behaviour (see [16]). A small peak can be observed within the central β (between 18 and 20 Hz) rhythm on the C4 spectral plots which suggests

that there is an ERS in the central β rhythm on ipsilateral hemisphere. The large peak on the μ rhythm on the C4 electrode is an electrophysiological correlate of cortical areas at rest or the cooperative or synchronised behaviour of a large number of neurons. Similar contralateral-ipsilateral differences occur during the imagination of left-hand movement, except the differences are symmetrically reversed. To determine that events are truly event related, an experiment described in [16], which involves averaging spectra, is the standard approach for distinguishing ERS/ERD in EEG signals.

The μ rhythm and the central β rhythm were selected as the most reactive frequency bands from which to extract features, for all subjects analysed. There are subtle differences in the main peaks of the upper and lower graphs in each figure indicating that throughout the imagination there is a change in the amplitude and degree of ERS/ERD in the signals. The evolution of the frequency over time within the FE window can be observed more closely by using an increased number of STFT windows with smaller length. Also, motor imagery data becomes most separable at specific segments (subject specific) [10], therefore if the FE window length M is selected properly the segments of data that produce maximum feature separability are captured as they pass through the window. The best FE window width M is subject specific. M must be selected empirically for each subject. If the STFT-window parameters are selected properly, then feature separability can be maximised within the FE window.

5. INTERPOLATION-BASED FEATURE EXTRACTION PROCEDURE

The extracted spectra contain quite a lot of detail on the frequencies that are not as prominent as those in and around the μ and central β ranges. Smoothing the spectra can reduce feature quality degradation caused by irregular frequency components introduced by noise and help compensate for missing information. The spectrum shape can be smoothed by decreasing the width of the STFT-window (i.e., increasing α of (1)). If the window is too narrow, the frequency resolution will be poor, and if the window is too wide, the time localization will not be so precise. This means that sharp localizations in time and frequency are mutually exclusive because a frequency cannot be calculated instantaneously [15]. Depending on the application and the quantity of information required about the frequency components the choice of window and window parameters must be adjusted to obtain the desired resolution. For this approach a good frequency resolution is important especially in the μ and β ranges but the objective is to obtain features which can provide maximum separability between both classes (left and right). In this respect the appearance of each spectrum was not of major importance. The reactive bands in the spectra are similar among most of the signals within each class but there are usually discrepancies in the upper and lower frequencies of each band as well as in the peak amplitude of each band. To reduce the possibility of these frequency components having a negative effect on the identification of features within each class an

interpolation process is performed to extract the gross shape of the spectrum.

The interpolation process can smooth the spectra and thus the differences between spectra within each class can be minimised. In this way some of the larger peaks may be lost but the interpolation plays a role in compensating missing information which ought to contribute to the discrimination [17] and can help reduce the intra class variance—a fundamental goal of most FEPs. The formula for the interpolation process is shown as follows:

$$Yip_l^k(u) = \sum_{i=\text{IPS}}^{\text{IPE}} \left(\frac{Y_k(i,l)}{\text{IPE} - \text{IPS} + 1} \right), \quad l = 1, \ldots, E, \quad (3)$$

$$\text{IPS} = \begin{cases} 0 & \text{if } u - ip < 0, \\ u - ip & \text{otherwise,} \end{cases} \quad (4)$$

$$\text{IPE} = \begin{cases} N_f - 1 & \text{if } u + ip > N_f + 1, \\ u + ip & \text{otherwise.} \end{cases} \quad (5)$$

Equation (2) is used to calculate Y_k, E is the number of spectra, and u is the value of the interpolated spectra at each frequency point (harmonic), therefore $u = 0, 1, \ldots, N_f - 1$. N_f is the number of frequency points or Fourier transforms in the spectra. The value of ip determines the number of interpolation points which in turn determines the degree of smoothing. A feature, f_l^k, is obtained by taking the l_2-norm (i.e., the square root of the sum of the components squared) of the interpolated spectra between the preselected reactive frequency bands. If E is the number of spectra (i.e., the number of windows for one spectrogram) and L is the number of signals, then

$$m = LE, \quad (6)$$

$$f_l^k = \left\| Yip_l^k \right\|, \quad l = 1, \ldots, E, \quad (7)$$

where f_l^k is a feature obtained from the reactive frequency bands of lth interpolated spectrum of the signal recorded at the kth recording site. According to (6), if there are 3 spectra (i.e., 3 STFT windows) within the FE-window for each signal, then $E = 3$, $L = 2$ (2 signals), and $m = 6$ thus, each feature vector would contain six features. To recapitulate, the number of features depends on the number of STFT-windows which depends on the FE-window length, the STFT-window length, and the amount of overlap between each STFT-window. If a large number of spectra are produced, choosing a number of specific interpolated spectra for feature extraction will reduce the feature vector dimensionality and thus maintain/improve computational efficiency; however this may cause performance degradation. The feature vector is

$$fv = (f_1^1, f_2^1, \ldots, f_E^1, f_1^2, f_2^2, \ldots, f_E^2). \quad (8)$$

6. CLASSIFICATION

After feature extraction classification is performed using linear discriminant analysis (LDA), a classifier that works on the assumption that different classes of features can be separated linearly. Linear classifiers are generally more robust than their nonlinear counterparts, since they have only limited flexibility (less free parameters to tune) and are less prone to overfitting [18]. Experimentation involved extraction and classification of features at every time point in a trial. The classes were labelled −1 for left and +1 for right. This resulted in a classifier which provides a time-varying signed distance (TSD) as described in [6, 11]. The sign of the classification indicates the class and the magnitude (or TSD) indicates the confidence in the classification. The time evolution of the CA rates and the TSD can be used to determine when the signals are most separable. The TSD is described in the following section.

7. PERFORMANCE QUANTIFICATION

The performance of the proposed BCI system is quantified by CA, IT rate, and the MI. The CA is the percentage of trials that are classified correctly. The capacity of a communication system is given by its IT rate, normally measured in bits/min (bpm). Capacity is often measured by the accuracy and the speed of the system in a specified application [19]. For systems that rely on accuracy and speed, the main objective is to maximise the number of bits that can be communicated with high accuracy in a specific time window. In present BCI systems, increasing the speed and accuracy is one of the main objectives. For example, the BCI systems in [4, 8, 19, 20, 21, 22] must be able to accurately decipher the EEG signals and respond correctly to its interpretation of the user's command as quickly as possible. IT rate was first used to quantify the performance of a BCI system by Wolpaw et al. [19] and the calculation was derived in [23, 24]. A relatively new quantifier of performance for a BCI system is to quantify the MI which is a measure of the average amount of information a classifier output contains about the signal. This performance measure was first used by Schlogl et al. [6, 7]. To estimate the MI the classifier should produce a distance value, D, where the sign of D indicates the class (in a two-class system) and the value expresses the distance to the separating hyperplane. A greater distance from the hyperplane indicates a higher signal-to-noise ratio (SNR). D is referred to as the time-varying signed distance (TSD) when estimated at the rate of the sampling interval. The D value at a specific time point t (i.e., $D(t)$) for all trials is used to estimate the MI. The MI between the TSD and the class relationship is the entropy difference of the TSD with and without the class information. The system described in this work facilitates features to be extracted with a time resolution as high as the sampling rate very easily, therefore the TSD is estimated at every time instant t although there must be M samples within the FE window before feature extraction begins.

8. SYSTEM OPTIMISATION

Due to the nature of this FEP, there are a number of parameters that must be tuned and the values of these parameters can have a significant effect on the performance of the system. These parameters are listed as follows:

(i) width of subject-specific frequency band(s),
(ii) FE window length, M,
(iii) STFT window length, N,
(iv) window width, α,
(v) overlap between STFT windows, ovl,
(vi) interpolation interval, ip.

Firstly, the most reactive frequency bands are selected. It is known from Pfurtscheller's work [16] and the Graz research group's [10] theoretical and meticulous work on EEG signals recorded during the imagination of left- and right-hand movement, as well as analysis done on the spectral graphs showing the ERD/ERS phenomenon for subject S1 (c.f. Section 4), that the most reactive bands usually occur in the μ (8–12 Hz) and central β (18–25 Hz) range. Further adjustments of the selected bands were carried out during the performance evaluation and it was observed that CA could be increased by adjusting the range of the selected bands. In this investigation an empirical selection of the most reactive frequency bands was performed by increasing or decreasing the μ and central beta bands in steps of 0.25 Hz. The data set for each subject was partitioned into three subsets—a training set (Tr), a validation set (V), and a testing set (Ts). The training sets consisted of 100 trials for subject S1, and 120 for subjects S2 and S3. The validation set for each subject consisted of 40 trials. The test (Ts) set consisted of 100 trials for subject S1, and 120 for subjects S2 and S3. The best subject-specific frequency bands and all other parameters were chosen by testing the system on a validation data set and choosing the band widths that provided the highest CA rates.

To begin the parameter selection procedure firstly, the FE window length, M, was chosen. The value of M had to be large enough so that the window contained enough signal to extract reliable features; however a window that is too large may result in degraded performance. For example, if a window length $M = 500$ is chosen, the minimum classification time is $500 \text{s} * 128^{-1} \text{s} = 3.9 \text{s}$ and if $M = 300$ the minimum classification time is 2.34 s therefore, the IT rate can be significantly influenced by the choice of M. Six different window sizes ranging between 100 and 450 were tested. The window size which provided the best features was selected for further tests. To tune the remaining STFT parameters firstly, 3 values of α were chosen and subsequently tests were run with $N = 50 : 50 : 300$ (i.e., N was set for all multiples of 50 up to 300) whilst ip and ovl were set to 1. It was assumed that by observing results at 6 different STFT window lengths, for each of the three different values of α, a sufficient indication of good combinations of these parameters for each subject could be attained. The highest CA rates on the training data were used to indicate the best combinations of all parameters. Up to eight different values of ovl were then selected ranging from 1 to 100 in specific multiples of 5 for small N

and 10 for larger values of N. The value of *ovl* must be less than N. At each value of *ovl* and the chosen best values of N and α, obtained from the first selection procedure, another set of tests were run with $ip = 3 : 3 : 18$. Again CA rates were used to choose the best combination of all four parameters. It was observed that the CA rates are sensitive to small changes in ip so another set of tests were carried out where the best chosen ip values from the previously described tests were decremented and incremented by 1 and then 2. In certain cases additional variations of the parameters were introduced for exhaustive tests. In a minority of situations the CA rates for two or more combinations were equal and in this case the IT rate was used to decide the best choice. This parameter selection technique only covers a small percentage of the possible combinations, therefore a more meticulous analysis may produce better results. An automated method could be used to search the parameter space for optimisation of the system.

9. RESULTS

All parameter selection was done by analysing how well the system performed on the validation data (40 trials for subject S1 and 60 trials for subjects S2 and S3). To test the generalisation abilities of the system, further tests were performed on the unseen testing data which consisted of 100 trials for each of the subjects. All performance quantifiers are estimated at the rate of the sampling interval (i.e., the performance is averaged over all trials at each time point; therefore, after each new sample is enveloped in the main FE window, the oldest sample is removed and a new set of features is extracted and classified). The results at the best time points (determined by the point at which CA is maximal) are presented. Table 1 shows the results obtained based on the parameters selected using the approach described in the previous section. Columns 1 and 2 indicate the subject and the selected subject specific frequency bands (2 frequency bands for each subject), respectively. There are three parameter combinations (PCs), and the corresponding results, shown for each subject. Column 3 specifies the PC for each subject for ease of reference. Columns 4–8 specify the FE window length M, the STFT window length N, the window width, α, the overlap between STFT windows, *ovl*, and the interpolation interval, ip, respectively. Column 8 specifies the number of features, m, which is calculated using (6). The CA rates for the validation data are specified in column 10. The CA rates, times at which CA is maximal (CT), the corresponding IT rates, and the maximum MI for the test data are specified in columns 11–14, respectively. All simulations were performed using MAT-LAB (http://www.mathworks.com). Functions from various toolboxes were utilised and all data manipulation and iterative software routines were developed using MATLAB source code.

9.1. Subject S1

From Table 1 it can be seen that the most reactive frequency bands and feature extraction parameters differ among subjects. For subject S1 the most reactive bands are within the entire μ range and a small band (18–19.5) within the central β range. When selecting the FE window size for subject S1, the CA rates for two different windows were equal; therefore, the STFT window parameters were selected for each of these windows and the results were compared. The best STFT window parameters differed for both FE windows. The CA rates on the test data were less than those achieved on the validation data indicating that overfitting occurred. PC2 achieved a higher CA rate on the validation data and also generalised the best to the test data. Also, the highest IT rates are not correlated with highest CA rates although the MI for PC2 is highest. As can be seen, the test CA rates for PC2 are only 1% higher than those obtained using PC1 but the IT rates are circa 3 bits/min lower—a substantial difference in IT rate. This is due to CT being much lower for PC1. The classification time is considered as the time interval (CT), beginning at the moment the user initiates the communication signal (i.e., second 3 of timing scheme [8]) and ending at the point where classification is performed. In an offline analysis, IT rate is calculated at the point where CA is maximal, thus providing an estimate of the maximum IT that the system is capable of achieving. The FE window size is significantly smaller for PC1 than for PC2 and, as mentioned in Section 4, this can affect the IT rate (i.e., the minimum CT is always $\geq M^*128^{-1}$). This is possibly the reason for significant differences in IT rates and indicates the importance of selecting the best FE window size.

9.2. Subject S2

The most reactive frequency bands for subject S2 were selected to be at the upper half of the μ band (10.75–13), the upper end of lower β band, and central β bands. In this case the CA rates of the test data are significantly higher than that of the validation data; however, the PC for this subject was chosen as the best and the results indicate that this PC generalises well to the test data. The difference in the CA rates may be due to the fact that the validation set is much smaller than the test set and may contain a larger percentage of trials which are more difficult to classify. The IT rate is significant at almost 9 bits/min. The MI for this subject is high, indicating that the SNR is high and that this subject may be able to perform modulated control of cursor more comfortably than subject S1.

9.3. Subject S3

The most reactive bands for subject S3 appeared to be between the upper end of the μ band and the lower end of the central β as well as in the upper β band. The upper β band is a fairly uncommon reactive band but the selection method described in Section 8 resulted in this band being chosen. For this subject the CA rates are, again, higher for the test data than for the validation data. This is possibly for the same reasons described for subject S2. The IT rate is significant at almost 12 bits/min. It can be seen that the CT is approximately 0.5 s less than that of subject S1 (PC2) but there is large difference in IT rates. This is due to the CT and the CA rates for each subject being substantially different. The MI for this subject is similar to that of subject S2.

TABLE 1: FEP parameter combination for three subjects and a comparative analysis of results.

| Sub | Parameters | | | | | | | No. | Val | Test | | | |
	2 freq. bands (Hz)	PC	M	N	α	ovl	ip	m	CA (%)	CA (%)	CT (s)	IT (b/m)	MI (bits)
S1	8–13,	1	200	50	0.68	1	4	8	90	85	2.18	10.7	.047
	18–19.5	2	360	100	3.68	15	2	8	91.3	86	3.42	7.28	0.52
S2	10.75–13, 17–22.5	1	360	100	0.68	1	4	6	86.3	91.7	4.11	8.56	0.65
S3	11.5–19.5, 27.25–30	1	360	50	0.68	5	4	14	87.5	91	2.98	11.33	0.63

10. DISCUSSION

10.1. System comparison

Results from this work show that the proposed FEP compares well to existing approaches. Performance results vary depending on different parameters choices. CA rates of 92% are achieved on unseen data without using cross-validation. Results ranging from 70% to 95% are reported for experiments carried out on similar EEG recordings [8, 9, 10]. Many of these results are subject specific and in some cases are based on a 10*10 cross-validation, results of which provide a more general view of the classification ability [8]. In [10] it is shown that the features derived from the power of the frequencies are most reliable for online feature extraction where results are obtained from 4 subjects, over a number of sessions. In the first few sessions the CA rates range between 73% and 85% and for later sessions the results range from 83% to 90%. The results in this work are based on recordings made in the first few sessions at early stages of training and results range between 85%–92%. Results are reported on tests across different sessions, indicating that the approach is fairly stable and robust for all subjects. Robustness appears to be an advantage of this approach, however an analysis for multiple subjects over multiple sessions is necessary to clarify this. Current BCIs have maximum IT rates of up to 25 bits/min [25]. In [26] it is shown that IT rates ranging between 12 and 18 bpm are achieved using left/right motor imagery data although, some of these results are based on a 10×10 fold cross-validation. In this investigation IT rates between 8–12 bits/min are achieved.

10.2. FEP parameters

Due to the considerably large number of possible FEP parameters combinations, all possible combinations were not tested. A more efficient way to find the optimum parameter settings would be to develop a fitness function which contains details on the three performance measures and the CT and use an automated search algorithm to optimise the PC. Criteria for limiting the optimisation to prevent over fitting may also be necessary. This would require a substantial amount of development and simulation time but would probably result in improved performance. For this analysis the results obtained were sufficient and compare well to results reported in BCI literature utilising similar data.

The selection of *subject-specific frequency bands* did significantly influence the results. The most reactive frequency bands were initially selected based on the visual inspection and then adjusted to obtain optimal performance (c.f. Section 8). In [10, 27, 28] the most reactive subject-specific frequency bands were selected by a technique known as distinction sensitive learning vector quantisation (DSLVQ) and it is shown that optimal electrode positions and frequency bands are strongly dependent on the subject and that subject-specific frequency component selection is very important in BCI systems. In [28] DSLVQ is applied on spectral data in 1 s time window starting after cue presentation whereas in this work the most reactive frequency bands were selected by analysing the time course of the CA rate. It is known that the frequency components may evolve during the course of the motor imagery tasks so it is possible that the most relevant bands vary during this period also. The empirical approach to frequency band selection employed in this work was used to find a general set of frequency bands for each subject so that CA could be maximised during the course of performing the mental task. Also, the bands were adjusted in steps of 0.25 Hz whereas in [28] the analysis was performed on 1 Hz bands ranging between 9 and 28 Hz. The approach carried out in this work was not overly time consuming and converged to a good set of relevant frequency bands for each subject. Although the approach described in this work is a manual approach, it may account for the evolving relevance of the frequency bands more so than the DSLVQ approach which is more automated but may have been more time consuming to perform an analysis such as that described in this work. In [28] it is suggested that, due to the relevance of frequency bands changing over the course of the trial, the DSLVQ algorithm may need dynamic adaptation to maintain optimal band selection. Future work will involve experimentation with DSLVQ to determine its potential for dynamically selecting the relevant frequency bands from EEG signals as they evolve during the course of the motor imagery tasks. This may enhance the accuracy and autonomy of the feature extraction procedure.

The *FE window length* can significantly influence the time course of the CA rates and CT. The best FE-window for all subjects appeared to be between 200–360 (i.e., between circa 1.56 s and 2.73 s long). None of the CTs equalled the window length, M, indicating that there was some data removed

(i.e., forgotten) from the FE window before data within the window became most separable. Therefore proper selection of the FE window can substantially improve performance by capturing only signal sequences which are most separable and forgetting data that may contribute to performance degradation.

The *STFT window parameters* (N, α, and ovl) are also crucially important for this approach. Most CA rates were maximised by using short but wide (small α) windows with small amounts of overlap. As detailed in Section 3, if the window is too narrow, the frequency resolution will be poor, and if the window is too wide, the time localisation will not be so precise. The temporal resolution can be made as high as possible by sliding the STFT window along the FE window with a large overlap. A small and wide STFT window ($M = 50$) can localise the frequency components in time whilst, at the same time, obtain a good frequency resolution. The window function utilised in this work becomes more like a uniform window with a parabolic top (i.e., less Gaussian) as α is decreased below 2. Therefore, most of the best PCs chosen cause the frequency components within each STFT window to be emphasised more so than a Gaussian window ($\alpha > 2$) would allow. The temporal resolution is achieved by sliding the STFT along the data with a certain overlap. Results from additional tests suggest that if the temporal resolution is too high (i.e., a large overlap) features overfitting may occur. N was set to 100 in the best PC for subject S2, indicating that the time localisation did not have to be as precise.

The *interpolation* process also plays an important role in the improvement of CA. The degree of smoothing is proportional to the value of ip. If ip is zero then no interpolation is performed. As can be seen from Table 1, for the best PCs for all subjects, some degree of smoothing was found to improve the CA rate. The improvement was, in some cases, only slight (approximately 2%) but nevertheless this is significant. The feature separability is very sensitive to the value of ip and increasing ip too much can cause performance degradation. As outlined in [19], a small increase in CA can significantly improve the IT rate, therefore the performance enhancement that the interpolation process can provide is very important in BCI systems. As mentioned, most of the PCs provided good time-frequency resolution but if the frequency resolution is too precise the intraclass variation will increase due to irregular frequency components. The interpolation process reduces the negative effects of irregular frequencies by smoothing the spectra and thus reducing the intraclass variance. Even increasing ip to 2 can reduce the intraclass variance and produce better CA and MI rates; however, in some cases, the interpolation process can reduce the interclass variance.

Overall, the parameters for each subject (apart from the subject-specific frequency bands and FE-window size) show some coherence. Therefore it may be appropriate to select a standard set for all subjects. This would allow fast application of the system to each individual subject. It is also possible that, by optimising the parameter combinations for each subject using an automated search algorithm, improved performance could be achieved, although the training times

may be costly. Parameters M and N do not have to be very finely tuned to obtain the best performance. Parameters α, ip, and ovl are critical parameters and cannot be varied too much from the selected best without significant degradation in performance. In additional experimentation, parameters were chosen arbitrarily with a small STFT window ($N = 50$) and high CA rates were achieved on the validation data but the results on the testing data were unsatisfactory. This occurred when ovl was large. For example, when the overlap was set equal to 45 (i.e., 95%), a large number of spectra were produced for each signal. Assuming the FE window size $M = 360$ then, the number of spectra (i.e., STFT windows) is $E = (M - ovl)/(N - ovl) = 63$ and from (6) $m = 126$ (cf. Section 3). This large number of features is almost half the number of data samples in the window and this can result in overfitted features. Thus the linear classifier begins to overfit. Parameter combinations that produced lower numbers of features (i.e., < 30) produced classifiers which generalised the best to the unseen test data.

10.3. The performance quantifying methods

The three performance measures have advantages and disadvantages and based on each, different conclusions can be drawn about the system. All three provide different information; classification accuracy rate simply provides the accuracy and other information such as sensitivity and specificity can be obtained. Even though these measures provide information about how well the system can distinguish between different sets of features extracted from the input space, they do not provide any information about the time required to do so. Timing is critical in any communication system and in most cases communicating in real time or as close as possible to real time is desirable. So, if a two-class system achieves 100% accuracy but it requires 20 seconds to perform the classification, then the advantage gained by the high accuracy is diminished by the fact that the classification required so much time.

As can be seen from Table 1, differences in CA and CT have significant effects on the IT rate, a performance measure which can quantify the performance of the system based on the CT and CA. The challenge is to find the optimal performance between accuracy and speed. In some cases the optimum can be obtained by accepting an FEP or classifier that has a reduced accuracy but a fairly rapid response. This will produce significantly faster IT rates but will result in a system where the probability of misclassifications occurring is much higher. This can be observed for the results of subject S1 where there is a slight difference in CA (1%) but a large difference in IT. The PC with the highest CA did not obtain the highest IT, therefore care must be taken when choosing the best PC.

MI calculation does not consider the accuracy or the time of classification but does quantify the average amount of information that can be obtained from the classifier output about the signal. This may be very important if it is intended to use the classifier output to control an application which requires proportional control. For example, the control of cursor may be performed by adjusting the cursor proportional

to the magnitude (TSD) of the classifier output and/or using the cursor to select from more than two choices on a one-dimensional scale. A person's ability to vary the MI would provide potential for the system to increase the possible IT rate to more than one bit for a two-class problem [6]. The MI can quantify how well a system may perform these types of tasks but does not provide much information about accuracy and time, therefore would not be a better quantifier than IT rate, although MI does provide information about the system that the IT rate does not. Overall maximising the CA rates is the most important although there is more useful information about the system performance contained in the IT rate.

11. CONCLUSION

To the best of the authors' knowledge, this type of TF-based FEP has not been used for feature extraction in EEG-based communication before. Although TF-based FEPs have been reported for application in BCIs, a process which involves a main FE window and interpolation process is a novel procedure and, as the results demonstrate, significantly enhances the FEP and overall system performance. Analysing the time evolution of the frequencies and values of the performance quantifiers can determine the best FE window size and also provides information about the signal segments which are most separable. The FE-window-based approach can be used for continuous feature extraction and thus has the potential to be used in an online system.

As the calculation of IT rate utilises knowledge on CA and duration of classification, IT rate provides significantly more knowledge about the system than simply the CA rate and the MI. However, classification accuracy is the most important in BCI applications and IT rates could be deceiving if CA and CT are not reported also. Therefore, it is concluded that, although IT rate is the best performance quantifier, all three quantifiers can provide information on different and important aspects of a BCI system. It is suggested that the results of each performance quantifier should be analysed and reported.

Further work will involve developing automated procedures for selecting the most reactive subject-specific frequency bands and an automated parameter optimisation procedure which can search the parameter space to find the optimum subject-specific parameters. Although, an empirical selection procedure can be used to select good subject-specific parameter combinations, it is anticipated that the full potential of the proposed approach will be realised only by developing a more intuitive parameter selection procedure.
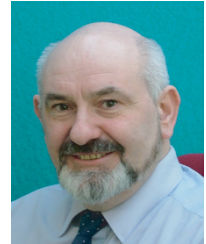
ACKNOWLEDGMENTS

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Journal of Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002, invited review.

[2] A. E. H. Emery, "Population frequencies of inherited neuromuscular diseases—a world survey," *Neuromuscular Disorders*, vol. 1, no. 1, pp. 19–29, 1991.

[3] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, et al., "Brain-computer interface technology: a review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[4] D. Coyle, G. Prasad, and T. M. McGinnity, "EEG-based communication: a time-series prediction approach," in *Proc. IEEE Cybernetics Intelligence—Challenges and Advances (CICA '03)*, pp. 142–147, Reading, UK, September 2003.

[5] B. J. Fisch, *Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG*, Elsevier, New York, NY, USA, 1999.

[6] A. Schlogl, C. Neuper, and G. Pfurtscheller, "Estimating the mutual information of an EEG-based Brain-Computer Interface," *Biomedizinische Technik*, vol. 47, no. 1-2, pp. 3–8, 2002.

[7] A. Schlogl, C. Keinrath, R. Scherer, and G. Pfurtscheller, "Information transfer of an EEG-based brain computer interface," in *Proc. 1st International IEEE EMBS Conference on Neural Engineering*, pp. 641–644, Capri Island, Italy, March 2003.

[8] C. Guger, A. Schlogl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller, "Rapid prototyping of an EEG-based brain-computer interface (BCI)," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 1, pp. 49–58, 2001.

[9] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent Neural Networks for EEG classification," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 457–463, 2000.

[10] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 3, pp. 316–325, 1998.

[11] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience Letters*, vol. 292, no. 3, pp. 211–214, 2000.

[12] A. Schlogl, D. Flotzinger, and G. Pfurtscheller, "Adaptive autoregressive modeling used for single-trial EEG classification," *Biomedizinische Technik*, vol. 42, no. 6, pp. 162–167, 1997.

[13] M. Roessgen, M. Deriche, and B. Boashash, "A comparative study of spectral estimation techniques for noisy nonstationary signals with application to EEG data," in *Proc. Conference Record of The 27th Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1157–1161, Pacific Grove, Calif, USA, November 1993.

[14] S. V. Notley and S. J. Elliott, "Efficient estimation of a time-varying dimension parameter and its application to EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 594–602, 2003.

[15] R. Q. Rodrigo, *Quantitative analysis of EEG signals: time-frequency methods and chaos theory*, Ph.D. thesis, Medical University of Lubeck, Lubeck, Germany, 1998.

[16] G. Pfurtscheller, *Electroencephalography, Basic Principles, Clinical Application and Related Fields*, E. Niedermeyer and F. L. Da Silva, Eds., Williams and Wilkins, Baltimore, Md, USA, 4th edition, 1998.

[17] D. Nishikawa, W. Yu, H. Yokoi, and Y. Kakazu, "On-line learning method for EMG prosthetic hand control," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 84, no. 10, pp. 35–46, 2001, *scripta technica*.

[18] K.-R. Muller, C. W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 165–169, 2003.

[19] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: improved accuracy by response verification," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 3, pp. 326–333, 1998.

[20] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 174–179, 2000.

[21] K. A. Moxon, "Brain-control interfaces for sensory and motor prosthetic devices," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 6, pp. 3445–3448, Salt Lake City, Utah, USA, May 2001.

[22] A. Kostov and M. Polak, "Parallel man-machine training in development of EEG-based cursor control," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 203–205, 2000.

[23] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Ill, USA, 1963.

[24] J. R. Pierce, *An Introduction to Information Theory*, Dover, New York, NY, USA, 1980.

[25] T. M. Vaughan, "Guest editorial brain-computer interface technology: a review of the second international meeting," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 94–109, 2003.

[26] P. Sykacek, S. Roberts, M. Stokes, E. Curran, M. Gibbs, and L. Pickup, "Probabilistic methods in BCI research," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 192–194, 2003.

[27] M. Pregenzer and G. Pfurtscheller, "Distinction sensitive learning vector quantization (DSLVQ) application as a classifier based feature selection method for a Brain Computer Interface," in *Proc. 4th International Conference on Artificial Neural Networks (ICANN '95)*, no. 409, pp. 433–436, Cambridge, UK, June 1995.

[28] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Trans. Rehab. Eng.*, vol. 7, no. 4, pp. 413–419, 1999.

Damien Coyle was born in 1980 in the Republic of Ireland. He graduated from the University of Ulster in 2002 with a First-Class Honours degree in electronics and computing engineering. He is currently undertaking research as a Ph.D. student in the Intelligent Systems Engineering Laboratory (ISEL) at the University of Ulster. His research interests include nonlinear signal processing, biomedical signal processing, chaos theory, information theory, and neural and adaptive systems. Coyle is a Member of the IEE and IEEE.
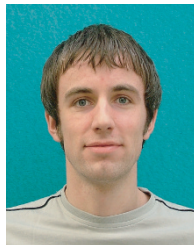
Girijesh Prasad was born in 1964 in India. He has a First-Class Honours degree in electrical engineering and a First-Class Master's degree in computer science and technology and received a Doctorate from the Queen's University of Belfast in 1997. Currently he holds the post of Lecturer and is a member of Intelligent Systems Engineering Laboratory (ISEL) research group of the School of Computing and Intelligent Systems at the University of Ulster. His research interests include computational intelligence, predictive modelling and control of complex nonlinear systems, performance monitoring and optimisation, thermal power plants, brain-computer interface, and medical packaging processes. He is a Member of the IEE and the IEEE and is a Chartered Engineer.

T. M. McGinnity has been a member of the University of Ulster academic staff since 1992, and holds the post of Professor of intelligent systems engineering within the Faculty of Engineering. He has a First-Class Honours degree in physics, and a Doctorate from the University of Durham, is a Fellow of the IEE, Member of the IEEE, and a Chartered Engineer. He has 25 years of experience in teaching and research in electronic engineering, leads the research activities of the Intelligent Systems Engineering Laboratory at the Magee campus of the university, and is Head of the School of Computing and Intelligent Systems. His current research interests relate to the creation of intelligent computational systems in general, particularly in relation to hardware and software implementations of neural networks, fuzzy systems, genetic algorithms, embedded intelligent systems utilizing re-configurable logic devices, and bio-inspired intelligent systems.