Feature-channel subset selection for optimising myoelectric human-machine interface design

Mohammadreza Asghari Oskoei*

Faculty of Mathematics and Computer Science,
University of Allameh Tabataba'i,
Tehran 15136-15411, Iran
and
School of Computer Science and Electronic Engineering,
University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK
E-mail: oskoei@atu.ac.ir
*Corresponding author

Huosheng Hu and John Q. Gan

School of Computer Science and Electronic Engineering,
University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK
E-mail: hhu@essex.ac.uk
E-mail: jqgan@essex.ac.uk

Abstract: This paper proposes a feature-channel subset selection method for obtaining an optimal subset of features and channels of myoelectric human-machine interface applied to classify upper limb's motions using multi-channel myoelectric signals. It employs a multi-objective genetic algorithm as a search strategy and either data separability index or classification rate as an objective function. A wide range of features in time, frequency, and time-scale domains are examined, and a modification that reduces the bias of cardinality in the separability index is evaluated. The proposed method produces a compact subset of features and channels, which results in high accuracy by linear classifiers without a need of preliminary tailor-made adjustments.

Keywords: myoelectric HMI; feature subset selection; multi-objective genetic algorithm; Davies-Bouldin index; DBI.

Reference to this paper should be made as follows: Oskoei, M.A., Hu, H. and Gan, J.Q. (2013) 'Feature-channel subset selection for optimising myoelectric human-machine interface design', *Int. J. Biomechatronics and Biomedical Robotics*, Vol. 2, Nos. 2/3/4, pp.195–208.

Biographical notes: Mohammadreza Asghari Oskoei is an Assistant Professor in University of Allameh Tabataba'i, Iran, and a Visiting Researcher in School of Computer Science and Electronic Engineering, University of Essex, UK. He received his MSc and BSc in Electrical Engineering from University of Tehran, Iran in 1993 and 1991, respectively. His received his PhD from School of Computer Science and Electronic Engineering in University of Essex, UK. His current research interests are in the areas of adaptive schemes in myoelectric control, machine learning, support vector machines with applications to biomedical signal processing and neuromuscular rehabilitation.

Huosheng Hu is a Professor in School of Computer Science and Electronic Engineering, University of Essex, UK, leading the Human Centred Robotics Group. He received his MSc in Industrial Automation from the Central South University, China in 1982, and PhD in Robotics from the University of Oxford, UK in 1993. His research interests include autonomous mobile robots, human-robot interaction, evolutionary robotics, multi-robot collaboration, embedded systems, pervasive computing, sensor integration, RoboCup, intelligent control and networked robotics. He has published over 350 papers in journals, books and conferences, and received several best paper awards. He is the Editor-in-Chief for *International Journal of automation and Computing*.

John Q. Gan is a Professor in School of Computer Science and Electronic Engineering, University of Essex, UK. He received his BSc in Electronic Engineering from Northwestern Polytechnic University, China in 1982, MSc in Automatic Control and PhD in Biomedical Electronics from Southeast University, China in 1985 and 1991, respectively. He has co-authored a book and published over 150 research papers. He is an Associate Editor for *IEEE Transactions on Systems, Man, and Cybernetics – Part B* and in the editorial boards of other journals. His research interests are in neurofuzzy computation and machine intelligence, brain-computer interfaces, robotics and intelligent systems, pattern recognition, signal processing, and data fusion.

1 Introduction

People with upper limb motor disability have problems with manipulating current assistive robots and rehabilitation devices that employ traditional user interfaces, such as a joystick, keyboard or keypad. They often either have lost their hand or have a paralysed hand that can not fulfil its motions properly; hence they need a comfortable, preferably wearable, hands-free human-machine interface (HMI), which requires minimal physical movements. Surface myoelectric signals (MES) collected from the skin of residual intact muscles contain rich motor control information that can be used to recognise muscular activities corresponding to various physical hand movements in a non-invasive manner. A myoelectric HMI uses MES as the reference input and has the potential to become an alternative to traditional body-powered user interfaces (Oskoei and Hu, 2007).

Pattern recognition-based myoelectric HMI functions by recognising pre-trained signal patterns and applying corresponding pre-defined commands. It provides hands-free and proximal HMI based on neuromuscular activities. The success of myoelectric HMI depends highly on classification rate. Applying a proper set of signal features and a suited classification method could enhance the classification rate. Meanwhile, involving proper channels of a multi-channel MES that provide most discriminating patterns is always a challenge that could directly affect the classification performance. It turns to a major challenge in the wearable devices that should provide an acceptable performance in various settings of electrodes on the limb. In this paper, we apply subset selection method to enhance the performance by optimising the structure of myoelectric HMI in terms of applied channels and features. The proposed method can be adopted for HMIs that use wearable high-density electrodes.

To collect effective signals, the electrodes of a myoelectric HMI should be placed over the limb concerning to its form, the desired motions and involved muscles. Therefore, the electrode placement typically requires help of experts who know limb anatomy and the desired motion kinematics. This is not always feasible and practical in regular daily-based applications. Subset selection methods could be applied to select distinctive channel of signals that provide high performance for the HMI. Thanks to active electrodes, there is no need of skin preparation and/or conductivity gel, hence, it is not difficult to use a relatively large number of electrodes in a wearable setup and adopt subset selection approaches to pick up a set of electrodes (i.e., channels) that provides the highest classification rate.

In general, raw data of MES are highly redundant and sometimes contradictory. Hence, their classification is time-consuming and inefficient. Features represent informative characteristics of the signals in a compact set, and their application results in relatively higher classification rate and presumably lower computational load (Oskoei and Hu, 2007). There are two distinct approaches to acquiring efficient features for classification, namely feature projection and feature selection. Feature projection creates a subset of new features by the combination of existing ones using linear or non-linear mapping functions, while feature selection chooses a subset of features by searching among the existing ones. Both approaches look for a feature set that comparatively results in high classification rate and low computational load. Dimensionality reduction is also a peripheral objective to both methods.

Englehart et al. (2001) applied feature projection to time-scale features and showed that principal component analysis (PCA) method provides a far more effective means of dimensionality reduction than feature selection. This was because time-scale methods produce a huge amount of signal features in both time and scale spaces, and dimensionality reduction becomes a significant challenge to simplify the structure of the classifier and meet real-time constraints. Chu et al. (2005) suggested a self-organising method along with PCA to create new feature space with good class separability. Most of the time-domain (TD) and spectral features are self-sufficient, and individually carry valuable information about muscular activities without curse of dimensionality. They only need to be sifted before applying to classification. Zardoshti-Kermani (1995) employed Davies-Bouldin index (DBI) and K-nearest neighbour classifier to select efficient MES features for classification. They examined some TD and spectral features, including integral average value, variance, zero crossings (ZCs), Wilson amplitude, v-order and log detectors, autoregressive model coefficients, and histogram. Park and Lee (1998) evaluated a set of MES features by comparing separability measure provided by the Bhattacharyya distance, and showed that the adaptive cepstrum vector (ACV) outperformed all other examined features. Chan et al. (2000) through developing a fuzzy classifier found out that the slope sign changes (SSC) feature introduced by Hudgins et al. (1993) did not improve classification performance or even deteriorated it.

The previous studies reveal the necessity of systematic procedures that are able to select efficient features for MES classification. Feature-channel subset selection (FSS) reduces computational load, improves classification performance and makes a better generalisation by turning the classifier to fewer parameters in pattern recognition. It also avoids surplus features that likewise the shortage can degrade classification performance (Handl and Knowles, 2006).

This paper proposes an evolutionary search algorithm to implement MES feature-channel subset selection for myoelectric HMI applied to upper limbs. In the proposed method, subset selection is added to offline system training procedure. The users can put on a wearable set of high density electrodes that cover most of the muscles on their limb and go through the system training procedure (i.e., feature-channel subset selection and offline training of the classifier) by presenting input MES corresponding to the desired motions (or muscular activates) and the desired classes of HMI outputs, respectively. In this study a wide range of features in time, frequency, and time-scale domains are individually and cumulatively evaluated. Furthermore, different selection criteria regarding the separability of features and their classification performance are examined.

The rest of the paper is organised as follows. Section 2 describes the proposed method for feature subset selection. Most known MES features are introduced in Section 3. Section 4 presents the experiments conducted to select the features as well as the channels having the highest contribution to classification performance. Discussion and conclusion are presented in Section 5.

2 Feature subset selection

Although FSS can be thought as a special case of feature projection, in practice it is a quite different problem. FSS is a directed search and looks at the issue of dimensionality reduction from different prospective and has a unique set of methodologies (Gutierrez-Osuna, 2005).

Given a set of features $F = \{f_1, f_2, \ldots, f_L\}$ the objective of FSS is to explore possible subsets $F_S \subset F$ with d_F features (i.e., cardinality of the feature set) that optimises an objective function $J(.)$.

$$F^* = \arg\min J\left(F_s\right) \tag{1}$$

FSS requires objective functions to evaluate the candidates and a search strategy to explore through all candidates. Objective functions are divided into two groups. The first evaluates candidates using their content (e.g., within and between clusters' separability, statistical correlations, information-theoretic measures), which is called the filter approach. The second group, known as the wrapper approach, applies classifiers to evaluate the candidates in terms of classification rate. Wrapper methods are very effective in reducing the feature space dimension and increasing classification rate. Their disadvantages include high computational load, susceptibility to overtraining, and the fact that their results are not often extendable to other type of classifiers.

Filter approaches typically select features based on their discriminative power. Popular methods in this respect include distance, dependency, information and consistency measures. As filter methods are independent of the classifier applied subsequently, they have good generalisation properties, but may be less effective in decreasing the dimensionality of the feature space and boosting classification rate. Generally, they are computationally cheaper than wrapper approaches (Handl and Knowles, 2006). Furthermore, filter methods that are based on distance computation in feature space suffer greatly a bias problem with respect to the cardinality of the feature set. The existence of this bias is related to the fact that, when moving to high dimensions, the histogram of distances between items in data space changes: the mean of the histogram tends to increase and the variance of the histogram tends to decrease. In other words, the distances between all pairs of points tend to become highly similar and this causes a bias. Consequently, they are biased towards low dimensions, and solutions in higher-dimensional space that are actually better than solutions in lower-dimensional space may be overlooked (Handl and Knowles, 2006).

The search space of finding d_F features out of L features is of size $\binom{d_f}{L}$. Although exhaustive search guarantees the global solution, it is unfeasible even for moderate values of d_F and L. Hence, a search strategy is needed to effectively direct the FSS process to explore the possible candidates, which adopts an objective function to evaluate the candidates and feedbacks a measure of 'goodness' (Gutierrez-Osuna, 2005). Possible choices for such search strategies include branch and bound, sequential search methods (such as forward selection, backward selection or floating search) and randomised search strategies (such as simulated annealing or evolutionary algorithms). An alternative and computationally cheaper approach is feature ranking, in which the individual features are evaluated in isolation only and those features with the highest capability are selected (Handl and Knowles, 2006).

In this paper, we employed individually DBI and Fishers linear discriminant index (FLD) as filter objective functions and linear discriminant analysis (LDA) and support vector machine (SVM) as wrapper objective functions, and developed a multi-objective genetic algorithm (GA) to explore MES features. To cope with the bias of cardinality, a modification suggested by Morita et al. (2003) was applied to the filter objective functions, and in order to make a compact feature set we considered the number of selected features (i.e., features cardinality) as the second objective function. This results in efficient and compact feature subset selection. Finally, an SVM classifier with either linear or non-linear kernels (Oskoei and Hu 2008) was individually employed to evaluate the performance of the selected subsets.

2.1 Objective functions

2.1.1 Davies-Bouldin index

This index represents separability of classes by the ratio of the sum of within-cluster distances to between-cluster distance (Bezdek and Pal 1998). To obtain the DBI, the similarity of each pairs of clusters that represent the level of distinction of two clusters should be computed. Clusters are formed by data samples labelled for certain activities in feature-channel space. Similarity of two clusters i and j is defined as

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{2}$$

where S_i and S_j are the dispersions of i^{th} and j^{th} clusters, respectively, and obtained by

$$S_i = \left\{ \frac{1}{|X_i|} \sum_{x \in X_i} \|x - m_i\|_2^q \right\}^{1/q} \tag{3}$$

where $|X_i|$ represents the number of samples in the i^{th} cluster and m_i is the mean defined as:

$$m_i = \frac{1}{|X_i|} \sum_{x \in X_i} x \ , \ \ x, m_i \in \Re^L \tag{4}$$

M_{ij} is the between-cluster distance defined by

$$M_{ij} = \left\{ \sum_{k=1}^{L} |m_{ik} - m_{jk}|^p \right\}^{1/p} \tag{5}$$

Assuming $p = q = 2$, we have adopted Euclidean distance as a measure between cluster's points and the mean. Let C be the number of clusters. DBI is defined as

$$DBI = \frac{1}{C} \sum_{j=1}^{C} \left(\max_{i \in C} R_{ij} \right) \tag{6}$$

As mentioned earlier, DBI is biased small feature subspaces. A possible approach to deal with this bias, suggested by Morita et al. (2003), is to normalise it to reverse the bias toward small feature subspaces. Hence, it is modified by:

$$DBI = \frac{1}{d_F} \frac{1}{C} \sum_{j=1}^{C} \left(\max_{i \in C} R_{ij} \right) \tag{7}$$

It is geometrically plausible to seek clusters that have minimum within-cluster distances and maximum between-class separation. For well-separated clusters, DBI is expected to decrease monotonically, and minimising DBI could be a reliable objective function that leads to the best subset of features to present MES for a classifier.

2.1.2 Fishers linear discriminate index

Developed originally by R.A. Fisher in 1936, it represents clusters' dispersion comparing to their scatter (Duda and Hart, 1973). It is defined as

$$FLDI = Trace\left(S_w S_b^{-1} \right) \tag{8}$$

where S_w and S_b are called within-class and between-class scatter matrixes, respectively. Let m be the mean of all clusters defined in (4). They are computed by

$$S_w = \sum_{i=1}^{C} \sum_{x \in X_i} (x - m_i)(x - m_i)^T \tag{9}$$

$$S_b = \sum_{i=1}^{C} |X_i| (m_i - m)(m_i - m)^T \tag{10}$$

Total scatter matrix is the sum of within-class and between-class scatter matrixes, which sometimes is applied instead of S_b in the FLDI computation:

$$S_t = S_b + S_w \tag{11}$$

Larger FLDI value shows higher possibility of linearly discriminating the clusters in the feature space. To keep uniformity with DBI we use the inverse of FLDI as the objective function that should be minimised.

$$iFLDI = \frac{1}{FLDI} \tag{12}$$

2.1.3 Linear discriminant analysis

It is a classic method of classification built on the same concepts introduced for FLDI. Given a set of samples of MES data belonging to different classes (i.e., a training set), LDA finds a good predictor that associates new samples (i.e., test set) into certain classes. The predictor is based on the mean and covariance of the samples in the training set for each class (Duda and Hart, 1973).

LDA often produces predictors whose accuracy approaches (and occasionally exceeds) more complex methods without any requisite to initial parameter adjustment. Hence, we employed LDA classifier as a wrapper objective function, and considered the rate of misclassification of the dataset represented by the desired features as an index that the search strategy seeks to minimise. In practice, LDA suffers from the singularity in covariance matrix of the feature sets, and this becomes a noticeable problem particularly in random search strategies.

2.1.4 Support vector machine

It is a powerful kernel-based classifier that constructs optimal boundaries between classes using the most informative samples in the training dataset (TDS). We preferred SVM with linear kernel to other kernels because it does not need any primitive parameter adjustment before application (Oskoei and Hu, 2008). The error of cross-validation, in which half of the data was for training and the rest for testing, was considered as an objective function that should be minimised in the FSS process.

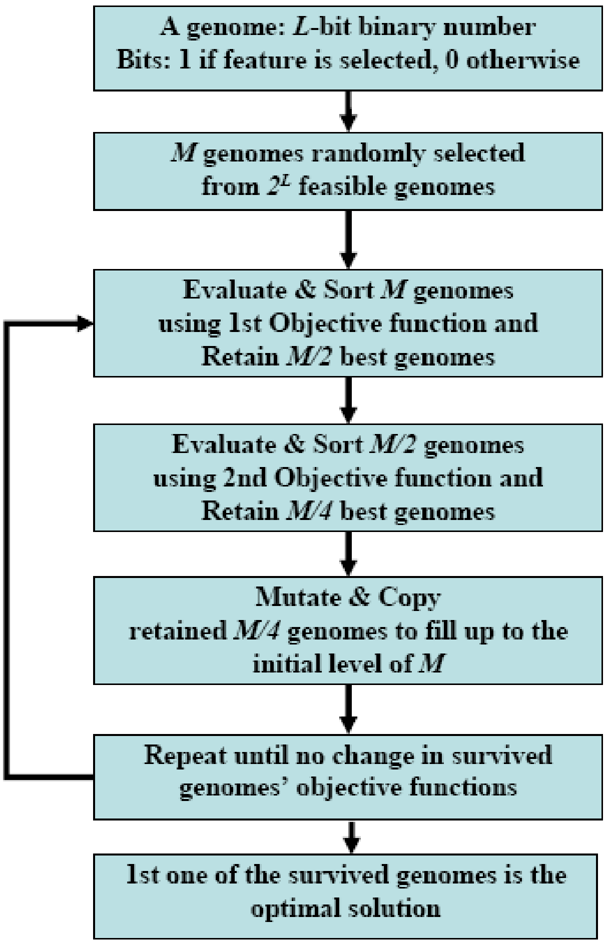
2.2 Search strategy

2.2.1 Genetic algorithm

It is an optimisation technique that mimics the evolutionary process of 'survival of the fittest'. Starting with an initial random population of solutions, it evolves new populations by mating (crossover) pairs of solutions and mutating solutions according to their fitness (objective function). The better solutions are more likely to be selected for mating and mutation, to carry their 'genetic code' from generation to generation.

Individual solutions, or the so-called genomes (i.e., candidate feature subsets), are simply represented with a binary string in which a bit is *1* if the given feature is selected, or *0* otherwise. Since GA yields several optimal solutions, the total number of selected features is considered as the second fitness criterion in a cascaded GA.

Figure 1 Flow chart of the cascaded GA applied as a search strategy in FSS (see online version for colours)



At the beginning, a population of 100 genomes (i.e., candidate subsets) is created. In each generation, all genomes are evaluated according to the first fitness criterion, and the 50 best individuals (the first elite) are retained while the rest are thrown away. The first elite are then sorted according to the second fitness measure and the best 25 genomes (the second elite) are retained. Mutated copies of the surviving genomes are then used to fill the population up to the initial level of 100. This multi-stage selection method enables the evolutionary algorithm to deal neatly with several fitness measures without the need to specify a joint fitness function, in which the relative contribution of each single factor (left to the insight of the designer) would play a very important role (Togelius et al., 2006; Eiben and Smith, 2003). Figure 1 demonstrates the flow chart of the proposed cascaded GA. In applying GA we should be very careful since it is shown that the performance of GA, though good for medium-sized problems, degrades as dimensionality increases.

3 MES features

This section presents briefly mathematical definitions of the features used for feature subset selection. These features belong to time, frequency, and time-scale domains. More detailed explanations can be found in Oskoei and Hu (2007, 2008).

3.1 TD features

TD features are computed based on instant signal amplitude and resultant values, giving a measure of waveform amplitude, frequency and duration within some limited parameters. Since they do not need a transformation, they are generally computed very quickly. Let x_i be the MES data samples, N the number of samples within a considered segment, and $x_{th} = 0.1 \times \arg\max_i |x_i|$ the amplitude threshold. Some popular TD features are defined as follows.

- mean absolute value (MAV):

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

- modified MAV (MAV1):

$$MAV1 = \frac{1}{N}\sum_{i=1}^{N} w_i |x_i| \ , \ w_i = \begin{cases} 1 & 0.25N \le i \le 0.75N \\ 0.5 & \text{else} \end{cases}$$

- modified MAV(MAV2):

$$MAV2 = \frac{1}{N}\sum_{i=1}^{N} w_i |x_i| \ ,$$

$$w_i = \begin{cases} 1 & 0.25N \le i \le 0.75N \\ 4i/N & 0.25N \ge i \\ 4(i-N)/N & 0.75N \le i \end{cases}$$

- mean absolute value slope (MAVS):

$$MAVS_k = MAV_{k+1} - MAV_k$$

- root mean square (RMS):

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$$

- variance (VAR):

$$VAR = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

- waveform length (WL):

$$WL = \sum_{i=1}^{N-1}\left|x_{i+1} - x_i\right|$$

- ZC:

$$ZC = \sum_{i=1}^{N-1} f_i$$

$$f_i = \begin{cases} 1 & x_i x_{i+1} < 0 \;\&\; \left|x_i - x_{i+1}\right| > x_{th} \\ 0 & \text{otherwise} \end{cases}$$

- SSC:

$$SSC = \sum_{i=2}^{N-1} f_i$$

$$f_i = \begin{cases} 1 & \begin{array}{l} x_i > x_{i-1}, x_i > x_{i+1} \,or\, x_i < x_{i-1}, x_i < x_{i+1}, \\ \left|x_i - x_{i+1}\right| > x_{th} \,or\, \left|x_i - x_{i-1}\right| > x_{th} \end{array} \\ 0 & \text{otherwise} \end{cases}$$

- Wilson amplitude (WAM):

$$WAM = \sum_{i=1}^{N-1} f\left(\left|x_i - x_{i+1}\right|\right),$$

$$f(x) = \begin{cases} 1 & x > x_{th} \\ 0 & \text{otherwise} \end{cases}.$$

Referring to Du and Vuskovic (2004) where windowing techniques (i.e., hamming and trapezoidal) were adopted for feature extraction, we have applied two windowing techniques, as shown in Figure 2, to calculate the modified versions of MAV, namely MAV1 and MAV2.

3.2 FD features

Frequency-domain (FD) features are calculated using either periodogram power spectrum density (PSD) or parametric methods. Compared to TD features, they require more computation power. Let x_i be the MES data samples, N the number of samples in a segment, and PSD be calculated through periodogram method and partitioned into M equal frequency bands with the average frequency f_i and corresponding average power P_i for each band. They are defined as follows.

- auto-regressive of n^{th} order (AR):

$$x_i = \sum_{j=1}^{n} a_j x_{i-j}$$

- PSD:

$$PSD = \frac{1}{M}\sum_{i=1}^{M} P_i$$

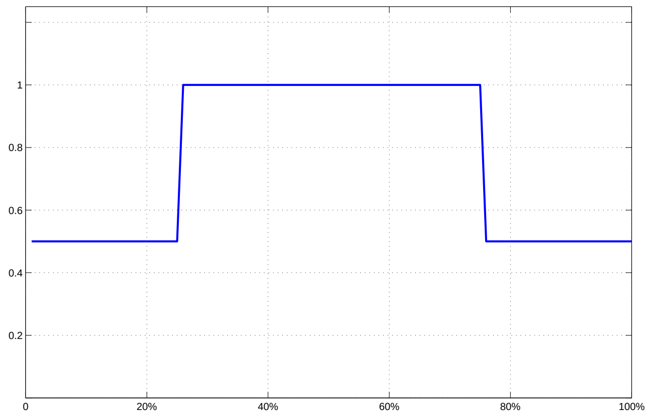
- mean frequency (MNF):

$$f_{mn} = \sum_{i=1}^{M} f_i P_i \Big/ \sum_{i=1}^{M} P_i$$

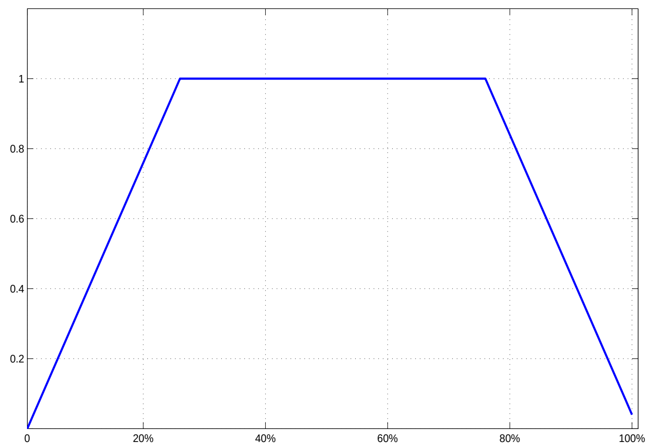
- median frequency (MDF):

$$\sum_{i=1}^{f_{md}} P_i = \sum_{i=f_{md}}^{M} P_i = \frac{1}{2}\sum_{i=1}^{M} P_i.$$

Figure 2 Modification applied on MES amplitude to calculate (a) MAV1 and (b) MAV2 features (see online version for colours)



(a)



(b)

With reference to Farina and Merletti (2000) AR coefficients with orders 2 and 6 have been considered in this work. They generate a 2- and 6-dimensional feature vector for each channel.

3.3 *Time-scale features*

Time-scale features are calculated by applying wavelet transform (WT) on raw MES data in each segment (Engelhart et al., 2001; Oskoei et al., 2008). The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled and shifted versions of the wavelet function. It results in coefficients that are functions of the scale and time shifts. Let CWT (s, τ) be the CWT of MES signals in scale s and time τ. The following feature is used in this work:

- Instantaneous mean frequency (IMNF):

$$IMNF = \frac{\int_0^{\tau 1}\int_0^{s1} |CWT(s,\tau)|^2 \, dsd\tau}{\int_0^{\tau 1}\int_0^{s1} s|CWT(s,\tau)|^2 \, dsd\tau}.$$

  Considering scales and times exclusively based on power two (i.e., dyadic scale and time) leads to the discrete wavelet transform (DWT) and the decomposition tool that decomposes MES signals into two parts: low-scale and high-scale. The low-scale part carries high-frequency components (i.e., the components of detail), while the high-scale part contains low-frequency components (i.e., the components of approximation). The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components (e.g., cD1, cD2, cD3 and cA3 as shown in Figure 3). This is called the wavelet decomposition tree. The sum of absolute values of these components individually represents the energy in different bands of scales (i.e., frequencies) and can be assumed as MES features:

- DWT components of nth order (*CDn, CAn*):

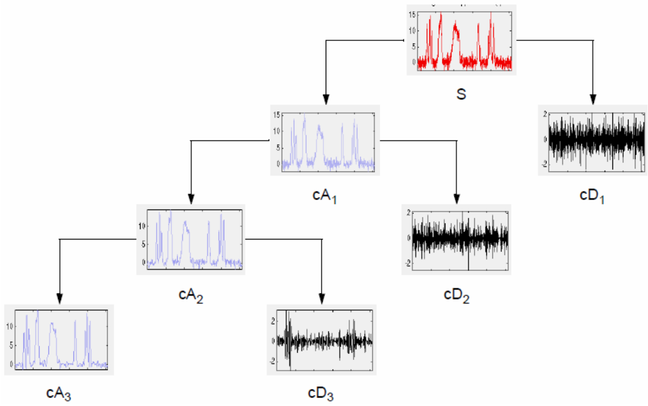
$$CD_n = \frac{1}{N}\sum_{i=1}^{N}|cD_n(i)|, \quad CA_n = \frac{1}{N}\sum_{i=1}^{N}|cA_n(i)|$$

  Three types of wavelets including *coif* (order 3), *db* (order 4), and *sym* (order 4) were applied to compute the IMNF and DWT components with order 4. Hence, they generate vectors with sizes of 1 and 5 for each channel, respectively. Time-scale features, similar to frequency domain features, impose a high load of computation.

3.4 *Group features*

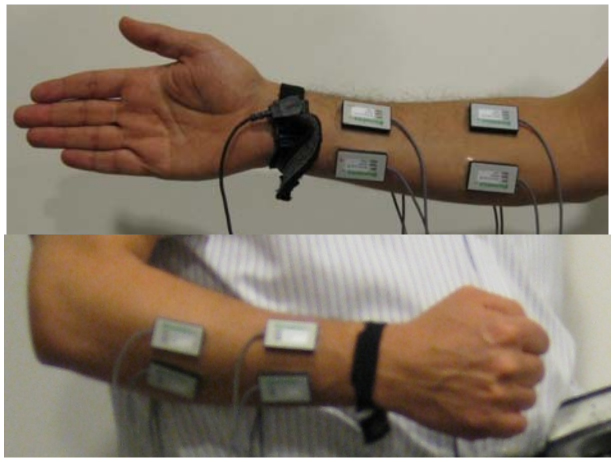
Four groups of features were examined in this paper. The first one named as TD feature set, was introduced by Hudgins et al. (1993) and includes MAV, WL, ZC, and SSC. The second group was recommended by Huang et al. (2005) and consists of RMS and AR6. The third and fourth groups are derivations of the first and second groups. The third one is formed by MAV, WL and ZC, while the fourth is constructed from RMS and AR2 (Oskoei and Hu, 2008). They produce features with dimensions of 4, 7, 3, and 3 for each channel, respectively. Hereafter, these groups are named as 'multi-feature'.

Figure 3 cD1, cD2, cD3, and cA3 are detail and approximation components of MES obtained by DWT decomposition with order three (see online version for colours)



Source: Misiti et al. (2006)

Figure 4 Electrode placement for the experiments (see online version for colours)



4 Experiments and results

This section describes the setup adopted to collect multi-channel MES corresponding to different hand states and then presents the results in three sub-sections: features evaluated individually, features selected by FSS, and channels selected by subset selection. Features and channels were examined individually in two experiments.

In the first experiment conducted to evaluate the features, a 4-channel MES was collected from two sides of a forearm (i.e., biarticulate wrist flexor, and triarticulate and biarticulate wrist extensor muscles) using bipolar active electrodes (Biometrics Ltd. SX230). An active electrode has a pre-amplifier with gain 1,000, which can differentiate between a small signal of interest and much larger interference signals that are present on the skin. It also has a very high input impedance to cope with mismatches in skin contact resistance. Signals are passed through a high-pass filter with a cut-off frequency of 10 Hz, to remove DC offsets due to membrane potentials, and to minimise interference due to electrode movement. A low-pass filter is used to remove unwanted frequencies above 450 Hz, and a notch filter used to remove unwanted line-frequencies (50/60 Hz). An electrode was also placed on the wrist, providing a common ground reference. Signals were sampled at 1,000 Hz using a 12-bit A/D converter (Figure 4).

Data were collected from eight healthy subjects in two sessions (16 sets). The subjects performed five limb motions plus resting, to produce six distinct states (i.e., classes). The motions were isotonic and comprised of flexion, extension, abduction, adduction, and keeping the hand straight. Two sequences of six motions, in which each motion was held fixed for five seconds, are called a block. Four blocks of data were gathered from subjects in each session. The collected data were segmented into adjacent windows with length of 200 ms and then features were extracted as described in the previous section.

4.1 Feature evaluation

To check validity of the applied measures, the features were individually evaluated in terms of their separability and resultant classification rate. The separability of features was calculated using DBI and FLDI defined in (6) and (12), individually. They represent the separability of the clusters that were formed in the feature spaces by MES data. Meanwhile, LDA and SVM classifiers were applied to evaluate the features in terms of classification rate. The classifiers were initially trained by a training set (random half of the collected MES data) and then applied to classify the test set (the rest of data). The error rate of classification was calculated using cross validation over the rate of misclassified samples in the test set. The result of evaluation is shown in Table 1. As shown, 25 sets of MES features including ten TD, five FD, six time-scale, and four group (multi-feature) features were evaluated. Dimension of the feature vector is also illustrated for each set. The figures in Table 1 depict the mean and standard deviation of the results obtained from sixteen independent datasets. They consist of the modified DBI, FLDI, LDA and SVM classification error rates. Low DBI or high FLDI indicates high rate of separability among the clusters, and low error rate of LDA or SVM depicts high potential of being classified by a linear classifier. Figures 5 to 8 illustrate the experimental results graphically.

Figure 5 Separability of MES features using modified DBI (see online version for colours)

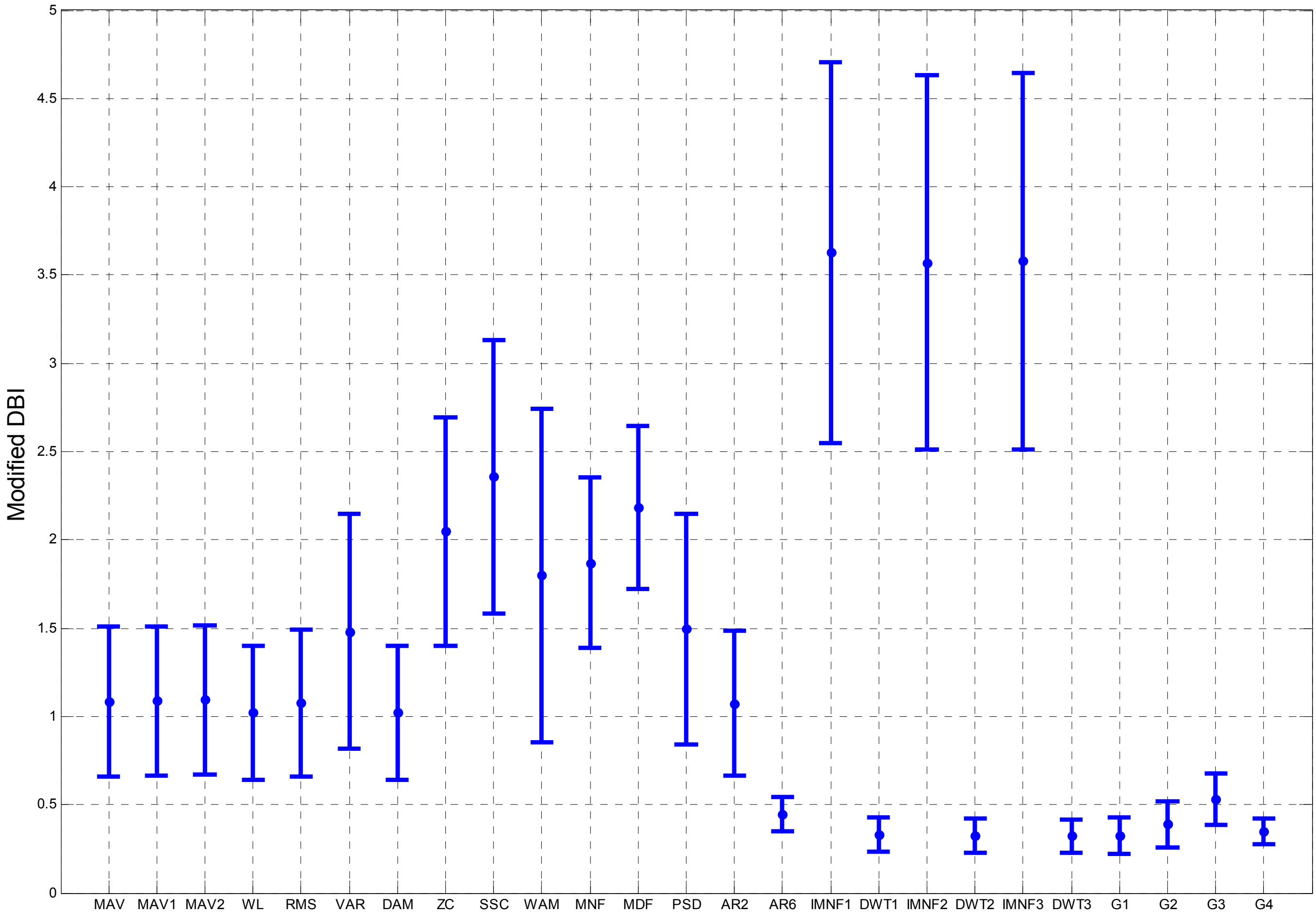


Figure 6 Separability of MES features using FLDI (see online version for colours)

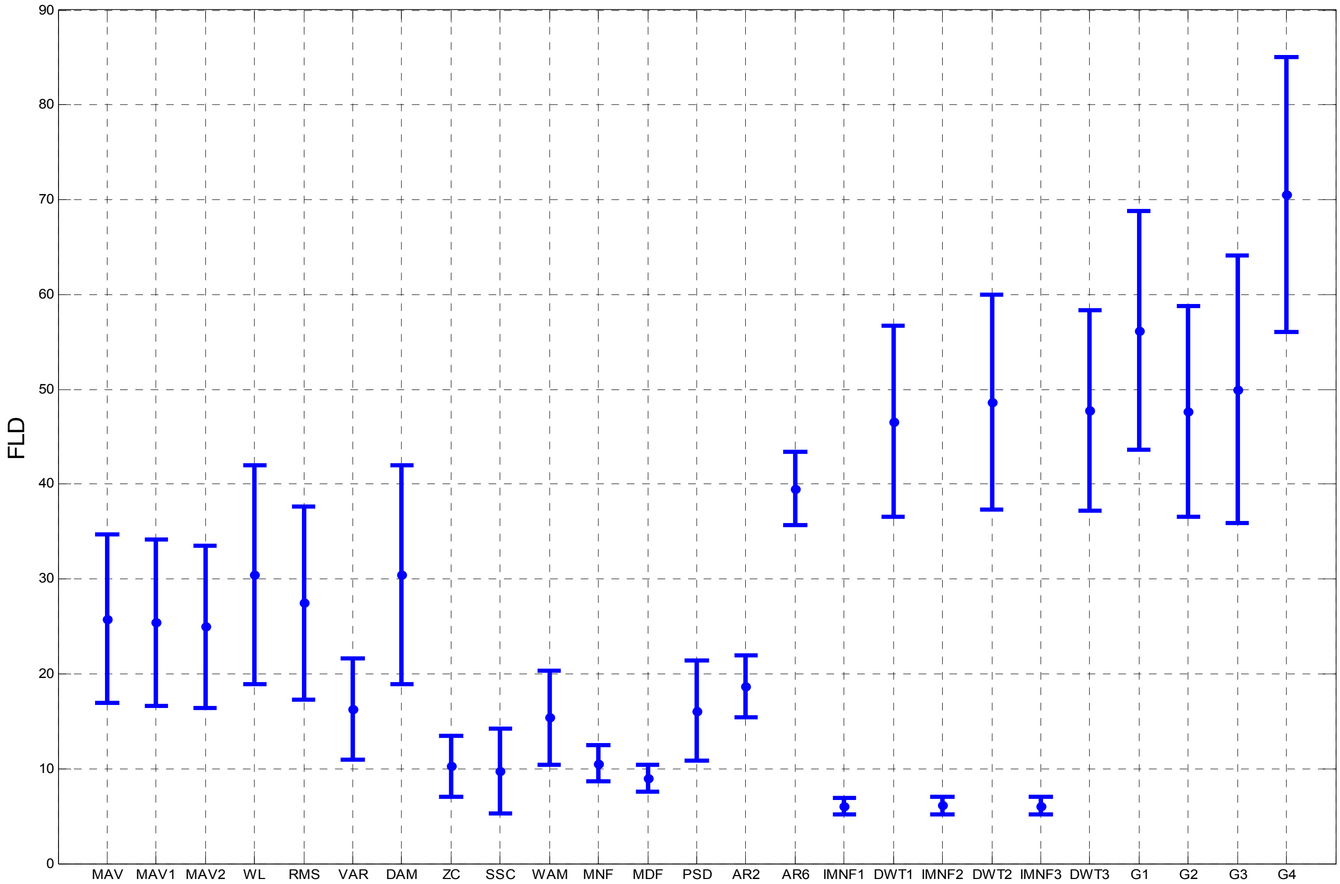


Figure 7 Classification error rate using LDA classifier (see online version for colours)

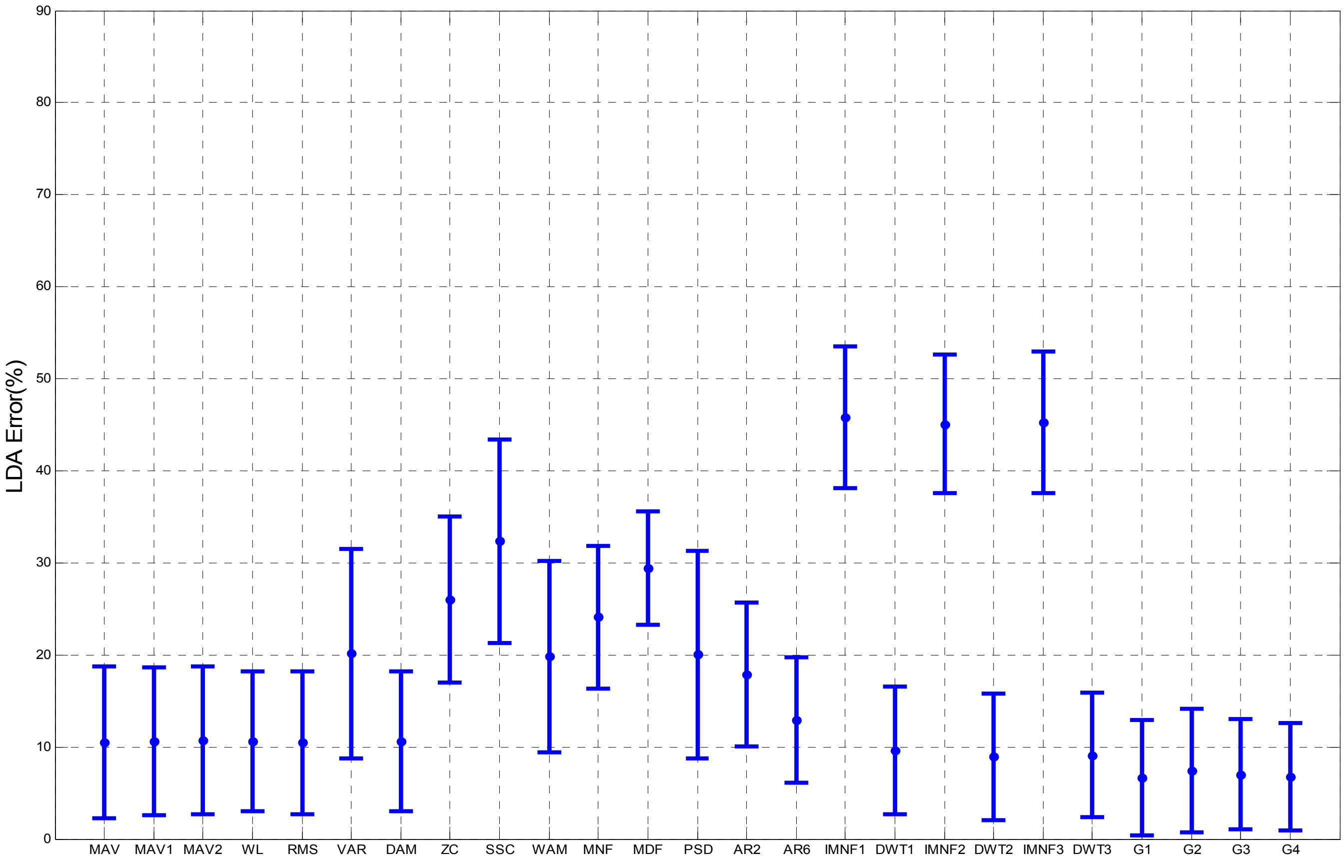
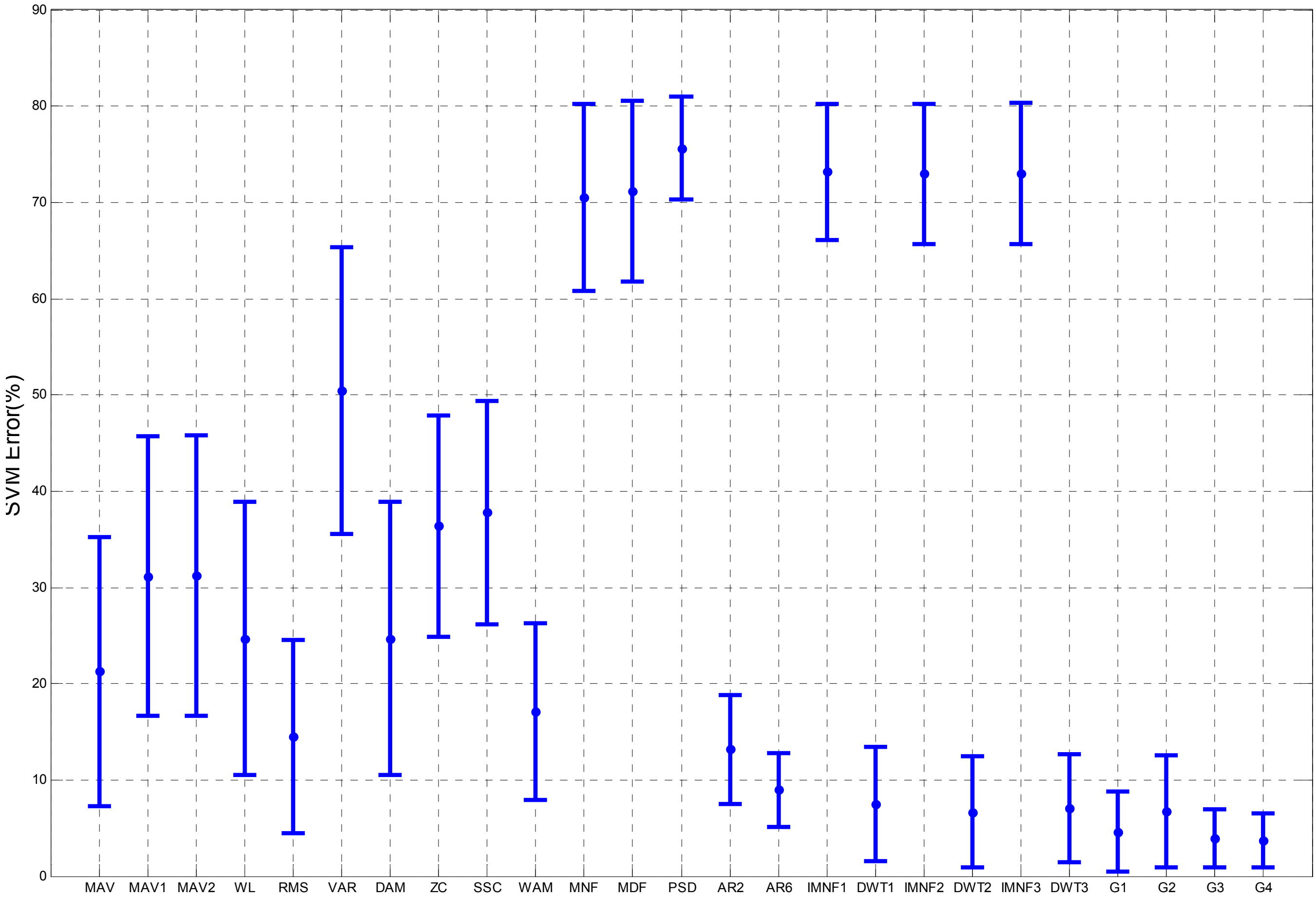


Figure 8 Classification error rate using SVM classifier (see online version for colours)



The results reveal that the multi-features (i.e., G1, G2, G3, and G4) generate higher rank in terms of both separability and classification rate. Among the single features, DWT, MAV, WL, RMS, and AR6 in order outperformed the others. Naturally, the computation of DWT and AR6 is more than that of MAV and WL, and the fact regarding to the slight difference in their performance makes the TD features attractive for real-time applications such as myoelectric HMI. MNF, MDF, and IMNF yield the lowest separability and classification hit rate, though, it is observed that they represent shift of frequency in long-term applications (Oskoei et al., 2008).

Furthermore, the findings in Table 1 can be used to study the 'bias of cardinality' in separability indexes. As mentioned earlier, DBI has a bias, and feature sets with higher dimensions are overlooked in comparison with low-dimensional subsets. The achieved results in Table 1 are consistent with this assumption. For instance, the DBI of AR6 feature (six-dimension) that yields higher classification rate is larger than AR2 feature (two-dimension) while rationally it is expected to be smaller. This phenomenon can also be observed in multi-features (e.g., G1) and DWT features. This approves using modified DBI instead of DBI itself in FSS as an objective function.

Moreover, the indexes in Table 1were examined in terms of consistency. Correlation coefficient was used to study the linear relationship between pairs of indexes. The closer coefficient to ±1 depicts the stronger correlation between the variables.

The findings reveal high consistency between the modified DBI and LDA classification error rate. The correlation coefficient of DBI and LDA error rate rises from 0.82 to 0.97 by applying the modification on DBI. It means that we can rely on the modified DBI to estimate the error rate of LDA classification. This fact is very significant in FSS, since LDA practically suffers singularity in covariance matrix of the data, which often happens due to randomness of the evolutionary algorithms in combing features and generating a new candidate. Meanwhile, FLDI and LDA classifier are both stemmed from a unique theoretical basis, and their relation is highly expected. The correlation coefficient between inverse of FLDI and LDA error is 0.99 and this confirms the mentioned assumption. However, it is degraded by dividing the FLDI by the features cardinality, hence, we just used inverse of FLDI without any modification regarding to the features cardinality.

Table 1 Evaluation of MES features individually using separability indexes and classification error rates

Features	*Dimension*	*Modified DBI*	*FLDI*	*LDA error (%)*	*SVM error (%)*
MAV	1 × 4	1.08 ± 0.42	25.82 ± 8.92	10.5 ± 8.2	21.3 + 14
MAV1	1 × 4	1.09 ± 0.42	25.42 ± 8.78	10.6 ± 8	31.2 + 14.5
MAV2	1 × 4	1.1 ± 0.42	24.99 ± 8.56	10.7 ± 8	31.2 + 14.6
WL	1 × 4	1.02 ± 0.38	30.42 ± 11.54	10.6 ± 7.6	24.7 + 14.2
RMS	1 × 4	1.08 ± 0.42	27.47 ± 10.15	10.5 ± 7.8	14.5 + 10.1
VAR	1 × 4	1.48 ± 0.67	16.31 ± 5.33	20.1 ± 11.4	50.5 + 14.9
DAM	1 × 4	1.02 ± 0.38	30.42 ± 11.54	10.6 ± 7.6	24.7 + 14.2
ZC	1 × 4	2.05 ± 0.65	10.28 ± 3.23	26.1 ± 9	36.4 + 11.5
SSC	1 × 4	2.36 ± 0.77	9.77 ± 4.49	32.4 ± 11.1	37.8 + 11.6
WAM	1 × 4	1.8 ± 0.94	15.4 ± 4.97	19.8 ± 10.4	17.1 + 9.2
MNF	1 × 4	1.87 ± 0.48	10.57 ± 1.87	24.1 ± 7.7	70.5 + 9.7
MDF	1 × 4	2.18 ± 0.46	8.99 ± 1.42	29.4 ± 6.1	71.2 + 9.4
PSD	1 × 4	1.5 ± 0.65	16.12 ± 5.31	20 ± 11.3	75.6 + 5.4
AR2	2 × 4	1.07 ± 0.41	18.67 ± 3.24	17.9 ± 7.8	13.2 + 5.7
AR6	6 × 4	0.45 ± 0.1	39.52 ± 3.87	12.9 ± 6.8	9 + 3.8
IMNF(coif3)	1 × 4	3.63 ± 1.08	6.03 ± 0.86	45.8 ± 7.7	73.2 + 7
DWT(coif3)	5 × 4	0.33 ± 0.1	46.62 ± 10.06	9.6 ± 6.9	7.5 + 5.9
IMNF(db4)	1 × 4	3.57 ± 1.06	6.11 ± 0.89	45.1 ± 7.5	73 + 7.3
DWT(db4)	5 × 4	0.32 ± 0.1	48.67 ± 11.32	8.9 ± 6.9	6.7 + 5.7
IMNF(sym4)	1 × 4	3.58 ± 1.07	6.1 ± 0.89	45.3 ± 7.7	73 + 7.3
DWT(sym4)	5 × 4	0.32 ± 0.1	47.77 ± 10.54	9.1 ± 6.8	7.1 + 5.6
MAV+WL+ZC+SSC (G1)	4 × 4	0.32 ± 0.1	56.22 ± 12.55	6.6 ± 6.3	4.6 + 4.1
MAV+WL+ZC (G2)	3 × 4	0.39 ± 0.13	47.7 ± 11.11	7.4 ± 6.7	6.7 + 5.8
RMS+AR2 (G3)	3 × 4	0.53 ± 0.15	49.98 ± 14.12	7 ± 6	3.9 + 3
RMS+AR6 (G4)	7 × 4	0.35 ± 0.07	70.59 ± 14.49	6.7 ± 5.8	3.7 + 2.8

Figure 9 Correlation between classification error rate and separability indexes after modification (see online version for colours)

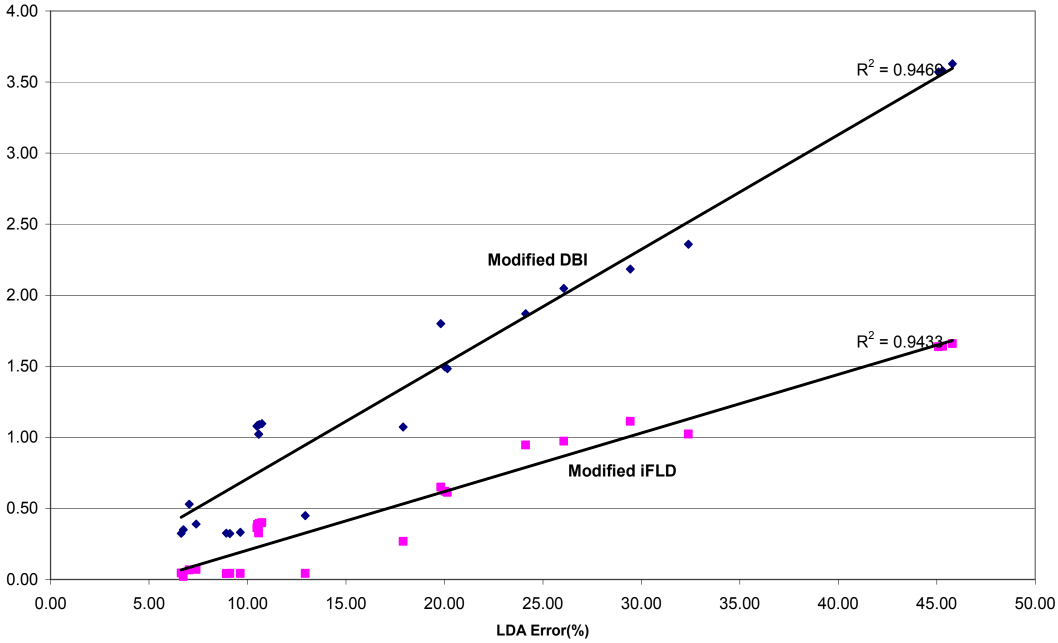


Figure 9 illustrates the scatter graph of classification error rate (wrapper objective function) and two separability indexes (filter objective functions) after modification. It depicts a linear regression composed of observations. The comparison of linear regressions shows that applying modification on DBI raises the R-squared value, which is a statistical term saying how good one thing is at predicting another.

4.2 Feature subset selection

We applied FSS with one of the four objective functions (i.e., modified DBI, inverse of FLDI, LDA and SVM), respectively, to find the best combination of the existing features. To make the selected subset as compact as possible, the number of features in each subset (its cardinality) was considered as the second objective function. Multi-objective GA was employed to handle concurrently two objective functions that should be minimised. Due to random initiation and mutation in the GA, the result of FSS is not unique and repeatable; hence, the mean of several iterations was considered for conclusion.

In practice, due to singular covariance matrix and/or between-clusters scatter matrix the FSS process using LDA or FLDI was problematic. GA randomly produces combinations that may cause a singular or very close to a singular matrix and this interrupts the search algorithm or makes it inaccurate. Therefore, results merely yielded by DBI and SVM were employed for conclusion. The selected features by FSS were fed into a classifier to evaluate their classification hit rate. We chose SVM with non-linear kernel (RBF) without any primitive parameter adjustment.

Table 2 Average number of selected features by FSS and their resultant classification rate without primitive parameter adjustment

Object function	*Dimension*	*Linear SVM (%)*	*Non-linear SVM (%)*
Normalised DBI	8 ± 1	98.7 ± 1.3	89.2 ± 7.5
Linear classifier	4 ± 1	98.9 ± 1.1	93.3 ± 6.2

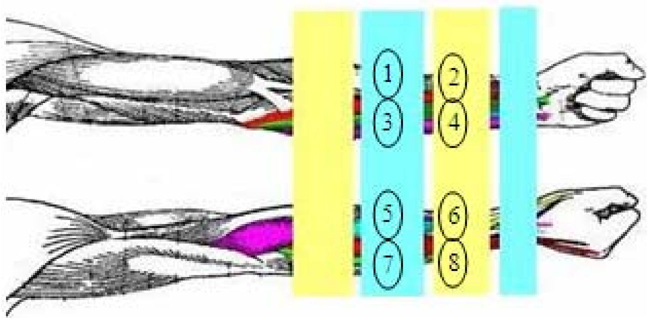
Half of the collected data was used as the TDS and the rest was used as the test dataset. The classification performance is obtained by cross-validation accuracy. The final result of fifteen iterations of FSS applied to sixteen independent datasets (i.e., totally 240 iterations) is shown in Table 2. It illustrates the average number of selected features along with their classification hit rate. The results indicate that the subsets selected by filter objective functions are much larger than those selected by wrappers, although both are capable of selecting subsets that result in high classification accuracy with a linear classifier and noticeable accuracy with a non-linear classifier. The FSS using wrapper objective functions outperformed the one using filters by resulting in more compact feature subsets and high rate of correct classification in both linear and non-linear boundaries. A glance at the selected subsets suggest that TD feature MAV, spectral feature AR2, and time-scale feature DWT (regardless to their wavelet function) were widely selected by FSS. Multi-features were also repeatedly selected by using both objective functions. It is seen that the bias of cardinality in DBI has got inversed and damped.

4.3 Muscle-channel selection

The objective of the second experiment is to select locations (i.e., muscles) on the forearm that provide the most distinct MES. Although this basically is a subject of physiology and kinesiology, processing multi-channel MES data corresponding to different hand states is an empirical approach that determines the required number of electrodes and their place (i.e., site) in such a way that HMI gains the best performance in terms of accuracy and response time.

There are three complex muscles in forearm, involved in hand motions through the wrist joint (ExRx.net LLC, 2006). They are known as wrist flexor, wrist extensor, and Brachioradialis. They contribute directly in various movements in joints around them such as wrist (e.g., flexion, extension, abduction, and adduction) and elbow (e.g., flexion and extension) as well as finger movements (e.g., flexion, extension) and thumb movements (e.g., flexion, extension, abduction, and adduction) (ExRx.net). They almost cover the entire two sides of forearm, and have several heads that function distinctly depending to the movements and joints.

Figure 10 Eight geometric parts of forearm to collect MES during different hand states (see online version for colours)



The exact locations of muscles' heads vary depending on forearm's anatomy. To avoid requisite of understanding the anatomy in electrode placement, we considered the forearm geometric shape and divided it into four parts in each side and then located totally eight bipolar electrodes in the centre of each part (Figure 10). All electrodes were placed in parallel to the muscles fibber and stuck to the skin. The MES data were collected during isometric contractions corresponding to six hand states (i.e., hand flexion, extension, abduction, adduction, kept-straight, and resting) from three subjects. TD feature set G3 (i.e., MAV, WL, and ZC) extracted from adjacent segments with length of 200 ms was used in this experiment.

Combinations of the channels were examined using the separability and classification rate of clusters that they generate. The LDA classification error rate versus the number of channels is shown in Figure 11. Both figures yielded by DBI and LDA classification error rate illustrate that 4-channel MES can produce a performance close to 8-channel MES performance. As expected, the results show that existence of channels from both sides of the forearm leads to higher rate of separability and classification. The statistics produced by applying FSS on the datasets indicates that a combination consisting of channels 2, 3, 6, and 7 builds clusters with high rate of separability. Figure 12 depicts the probability of selection of MES channels during five iterations of FSS applied on three collected datasets (i.e., totally 15 iterations).

Figure 11 Classification error rate vs. number MES channels (see online version for colours)

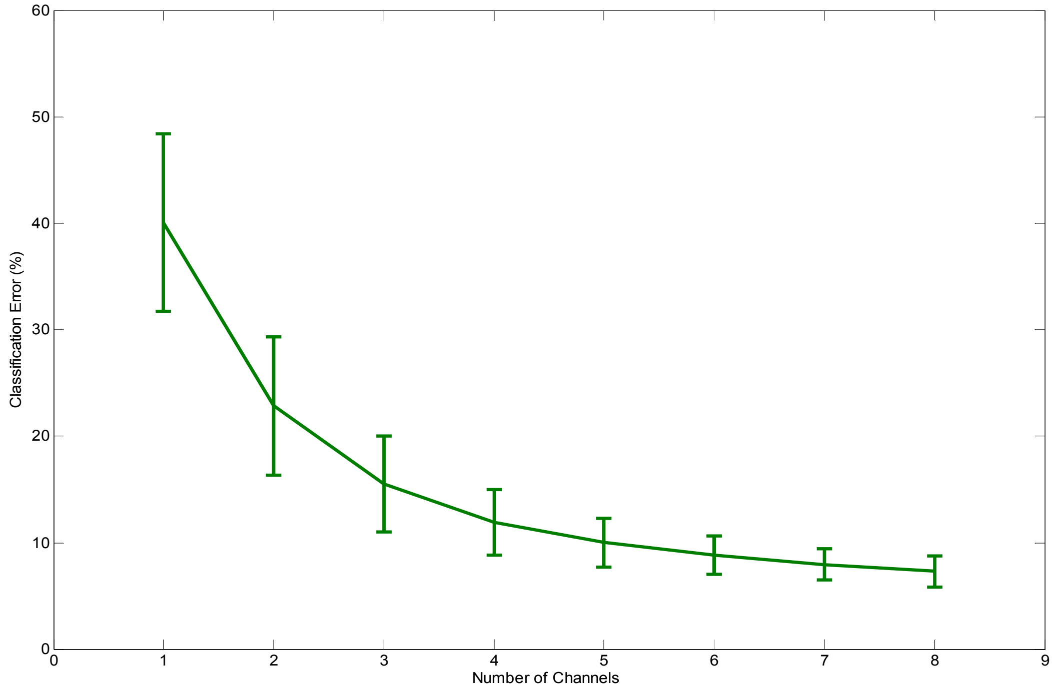
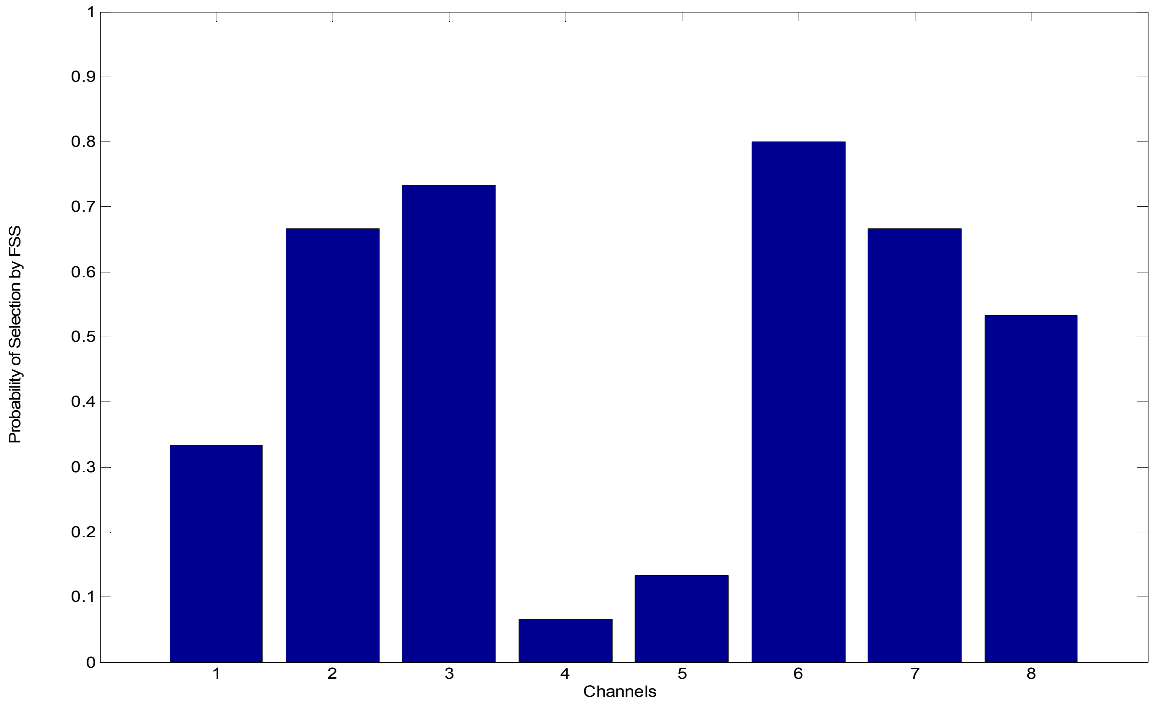


Figure 12 Distribution of selected channels using modified DBI (see online version for colours)



5 Conclusions

The proposed feature-channel subset selection method was applied to design a pattern recognition-based myoelectric HMI, which provided an automatic routine to configure the HMI during offline training and made the user dispense with tailor-made adjustments. The proposed FSS applies evolutionary methods to explore feasible subsets of MES features and employs data clusters' separability and classification rate to evaluate features. It has been shown that the selected feature subsets result in high classification accuracy through a linear classifier. Filter objective functions result in larger feature subsets in comparison with wrapper objective functions. Normalisation reduces the bias of cardinality in DBI, while it is unnecessary for FLDI. Meanwhile, the evaluation of a wide range of MES features shows that TD multi-features, because of their simplicity and light computation load and just a slightly lower classification rate compared to time-scale and spectral features, are the best option for a real-time HMI. Finally, it is found out that 4-channel MES collected from both sides of a forearm provides enough information to recognise muscular activities corresponding to different hand states. Future work will focus on the development of unsupervised methods for feature-channel subset selection to achieve accurate classification of upper limb motions using MES.

References

Bezdek, J. and Pal, N. (1998) 'Some new indexes of cluster validity', *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, Vol. 28, No. 3, pp.301–315.

Chan, F., Yang, Y., Lam, F., Zhang, Y. and Parker, P. (2000) 'Fuzzy EMG classification for prosthesis control', *IEEE Trans. on Rehabilitation Eng.*, Vol. 8, No. 3, pp.305–311.

Chu, J., Moon, I., Kim, S. and Mun, M. (2005) 'Control of multifunction myoelectric hand using a real time EMG pattern recognition', International *IEEE Conference on Intelligent Robots and Systems*, pp.3957–3962.

Du, S. and Vuskovic, M. (2004) 'Temporal vs. spectral approach to feature extraction from prehensile EMG signals', *International IEEE Conference on Information Reuse and Integration*, pp.344–350.

Duda, R. and Hart, P. (1973) *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, NY, ISBN: 9780471223610.

Eiben, A. and Smith, J. (2003) *Introduction to Evolutionary Computation*, Springer-Verlag, Berlin Heidelberg, Germany.

Englehart, K., Hudgins, B. and Parker, P.A. (2001) 'A wavelet-based continuous classification scheme for multifunction myoelectric control', *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 3, pp.302–310.

ExRx.net LLC (2006) 'ACSM resource manual for guidelines for exercise testing and prescription' [online] http://www.exrx.net/Lists/ExList/ForeArmWt.html, pp.224–349.

Farina, D. and Merletti, R. (2000) 'Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions', *Journal of Electromyography and Kinesiology*, Vol. 10, No. 5, pp.337–349.

Gutierrez-Osuna, R. (2005) *Introduction to Pattern Analysis*, Lecture Notes, Texas University.

Handl J. and Knowles, J. (2006) 'Feature subset selection in unsupervised learning via multiobjective optimization', *International Journal of Computational Intelligence Research*, Vol. 2, No. 3, pp.217–238.

Huang, Y., Englehart, K., Hudgins, B. and Chan, A. (2005) 'A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses', *IEEE Trans. on Biomedical Engineering*, Vol. 52, No. 11, pp.1801–1811.

Hudgins, B., Parker, P. and Scott, R. (1993) 'A new strategy for multifunction myoelectric control', *IEEE Transaction on Biomedical Engineering*, Vol. 40, No. 1, pp.82–94.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J. (2006) *Wavelet Toolbox User's Guide*, MathWorks, Inc., Natick, MA, USA.

Morita, M., Sabourin, R., Bortolozzi, F. and Suen, C. (2003) 'Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition', *International Conference on Document Analysis and Recognition*, IEEE Press, pp.666–671.

Oskoei, M.A. and Hu, H. (2007) 'Myoelectric control systems – a survey', *Biomedical Signal Processing and Control* (Elsevier), Vol. 2, No. 4, pp.275–294.

Oskoei, M.A. and Hu, H. (2008) 'Support vector machine-based classification scheme for myoelectric control applied to upper limb', *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 8, pp.1956–1965.

Oskoei, M.A., Hu, H. and Gan, J. (2008) 'Manifestation of fatigue in myoelectric signals of dynamic contractions produced during playing PC games', *International IEEE Conference of Engineering on Medicine and Biology Society*, pp.315–318.

Park, S.H. and Lee, S.P. (1998) 'EMG pattern recognition based on artificial intelligence techniques', *IEEE Transactions on Rehabilitation Engineering*, Vol. 6, No. 4, pp.400–405.

Togelius, J., De Nardi, R. and Lucas, S. (2006) 'Making racing fun through player modeling and track evolution', *Proceedings of the SAB Workshop on Adaptive Approaches to Optimizing Player Satisfaction*.

Zardoshti-Kermani, M. (1995) 'EMG feature evaluation for movement control of upper extremity prostheses', *IEEE Trans. on Neural Systems & Rehabilitation*, Vol. 3, No. 4, pp.324–333.