View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Ulster University's Research Portal

7

Recent Advances in Prediction-based EEG Preprocessing for Improved Brain-Computer Interface Performance

Damien Coyle
Intelligent Systems Research Centre, University of Ulster
Northern Ireland, UK

1. Introduction

Brain-computer interface (BCI) technology is an assistive and augmentative technology that has the potential to significantly enhance the quality of the lives of those who require an alternative means of communicating and interacting with people and their environment. BCI research is growing at a significant pace (Vaughan and Wolpaw, 2006; Wolpaw et al., 2002; Mason et al., 2007; Lecuyer at al., 2008; McFarland and Wolpaw, 2008; Coyle et al., 2005a, 2006a) with many advances in signal processing and a range of BCI applications being investigated in the past few years. The depth and breadth of BCI research in progress today is indicative of its application potential – this is exemplified by the year-on-year exponential increase in peer review journal publications, regular news items in the media, formation of BCI related companies and substantial investment in BCI-specific projects. Being able to offer people with limited neuromuscular control, due to disease, spinal cord injury or brain damage (Wolpaw et al., 2002) an alternative means of communication through BCI will have an obvious impact on their quality of life. A range of studies have shown that head trauma victims diagnosed as being in a persistent vegetative state (PVS) and locked-in patients due to motor neuron disease or brainstem stroke may specifically benefit from BCI systems (Wolpaw et al., 2002; Mason et al., 2007; Owen and Coleman, 2008; Silvoni et al., 2009; Birbaumir et al., 1999; Kaiser et al., 2001) although, as BCIs improve and surpass existing assistive technologies, they will be beneficial to those with less severe disabilities (Pfurtscheller et al., 2007) and applications such as neurofeedback for stroke rehabilitation (Prasad et al., 2009), epileptic seizure prediction (Iasemidis, 2003), driver awareness/alertness detection and cognitive load monitoring. BCI is also emerging as an augmentative technology in computer games (Lecuyer at al., 2008), virtual reality (Leeb et al., 2007) and robotics (McFarland and Wolpaw, 2008).

Even though BCI technology has been under investigation concertedly for the past ten years (Vaughan and Wolpaw, 2006; Mason et al., 2007), there remain many challenges and barriers to providing this technology easily and effectively to the intended beneficiaries. These challenges include i) identification of the most appropriate mental tasks and EEG signals; ii) enhancing training through better feedback and reduced training durations; iii) developing hardware for ambulatory EEG – unobtrusive, practical, low power consumption and cost effective; iv) developing better biosignal processing algorithms (preprocessing, feature extraction/selection/translation, classification and post-processing) to improve performance (classification accuracy (CA), information transfer (IT) rates and reliability; v) enabling long-term and short-term autonomous system adaptability; vi) developing BCI-specific intelligent applications; and vii) assessing user acceptance and the service and care required at the initial stages (Wolpaw et al., 2002).

There have been significant advances in addressing these issues, but often, whilst one issue is addressed another arises. For example, it is often the case that using more electrode channels in a motor imagery based BCI provides better performance than a BCI with less channels – due to a better spatial resolution and the identification of subject-specific cortical activity topography. However increased electrodes significantly reduce the practicality of the BCI and increase the obtrusiveness of the montage. Other issues arise with large montages because the best currently available electrodes require electrolyte gels which can be messy and time consuming to apply, although dry electrodes are available but not widely used as yet (Popsecu et al., 2007). Another example of how improvements in one aspect of a BCI can have implications for other aspects is the subject-specific hyperparameter tuning problem. Almost all signal processing methods can be improved by tuning hyperparameters and tailoring signal processing methods specifically to each subject, sometimes referred to as calibrating the system. In many cases this is done offline manually or semi-automatically with heuristic approaches using data obtained via a training session. This is an effective approach and often considered essential however it does pose challenges for offering BCI widely to multiple individuals where minimal parameter tuning and operator interaction is required. BCIs require signal processing algorithm that can be applied and adapted easily and online automatically to accommodate user adaptation and drifts in attention, mood and fatigue levels. A BCI which does not require extensive parameter tuning and tightly bounded parameters but a more general set of parameters may be able to accommodate better accuracies and robustness in the face of such changes and may be more conducive to autonomous adaptation where only generalized changes to a minimal number of parameters are necessary.

A range of studies have been undertaken to address these issues but the main emphasis in BCI is on enhancing the separability of features extracted from EEG signals associated with various brain states and using advanced classification techniques to maximize the accuracy in classifying those brain states. For example, the neural-time-series-predication-preprocessing (NTSPP) framework increases data separability by predictive filtering and mapping the original EEG signals to a higher dimensional space using predictive/regression models which have been individually specialised (trained) on EEG signals associated with specific brain states (Coyle et al., 2004; 2005a; 2006a; 2006b; 2008a; 2009). Features extracted from the mapped space are more separable than those produced by the original EEG signals, in terms of increased Euclidean distance between class means and reduced inter-class correlation and intra-class variance. Preliminary results from recent work (Coyle et al., 2008a) show that NTSPP compares well to the spatial filtering approach known as common spatial patterns (CSP) (Blankertz et al., 2008; Dornhege et al., 2006; Ramouser et al., 2000) which is used extensively in BCI research. The results also indicate that CSP can complement NTSPP using a reduced electrode montage with no subject-specific parameters; producing a 3-channel BCI that achieves performance which is comparable to a 60 channel BCI in certain cases when no subject-specific parameter tuning is carried out (Coyle et al., 2008a). CSP constructs linear spatial filters that maximize the ratio of class-conditional variances of EEG sources (Ramouser et al., 2000) and can also be used to reduce the dimensionality of the feature vector by providing a surrogate data space with less data. When NTSPP is employed in a 2-class, multichannel system the data dimensionality can increase significantly whereas CSP can reduce the dimensionality of a multidimensional signal space, and both can improve separability, therefore the NTSPP-CSP combination offers significant potential for improved and stable performance in BCI systems. Additionally, it has been shown that using subject-specific discriminable frequency bands or spectral filtering (SF) improves overall BCI performance. Spectral features of the EEG are widely used in MI-based BCIs because lateralized neuronal activity in motor cortical areas is usually distinguishable in mu (8-12Hz) and central beta (18-25Hz) frequency bands (Blankertz et al., 2008; Pfurtscheller et al., 1998; Pfurtscheller, 1998; Coyle et al, 2005b; Herman et al., 2008). In addition to NTSPP and CSP, subject-specific SF can be employed, resulting in a temporal-spectral-spatio preprocessing framework (NTSPP-SF-CSP).

Developing approaches which can address all signal processing related issues is a challenge however the hypothesis of this work is that the neural-time-series-prediction-preprocessing (NTSPP) framework offers the potential of making BCI simpler (negating the need for subject-specific hyperparameters and minimizing the number of electrode channels required) whilst maintaining or enhancing performance of existing BCI methods. The aim of this chapter is to present a comprehensive analysis of NTSPP and its capacity to address a number of the issues in BCI, as outlined above, and to determine the advantages of employing multiple EEG channels in a 2 class motor imagery BCI (22 channels) compared to 2 and 3 channel montages. To achieve these aims data from twenty-three BCI subjects are used and the analysis carried out has the following objectives.

1. to compare the performance differences between BCIs employing spectral filtering (SF) only, SF and CSP combined (SF-CSP), NTSPP-SF combined, and NTSPP-SF-CSP combined.
2. to show that NTSPP can complement CSP using a reduced electrode montage with minimal subject-specific parameters.
3. to compare performances with 2 electrodes, 3 electrodes and 22 electrodes all with standard positioning.

Also, to conduct a fairer comparison[1] of all methods, a range of different classifiers have been investigated including various statistical classifiers such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and other distance based classifiers all of which are available in the Biosig tool box (Schlogl, 2007). A probabilistic Bayes based classification method with evidence accumulation is also tested in addition to a committee based approach involving all classifiers are also tested.

The chapter is structured as follows. Section 2 provides information on the datasets used and the data acquisition process. Section 3 describes the methods employed including NTSPP and the self-organizing fuzzy neural network (SOFNN) which is used in the NTSPP framework. CSP and feature extraction methods and a brief description of the classifier and

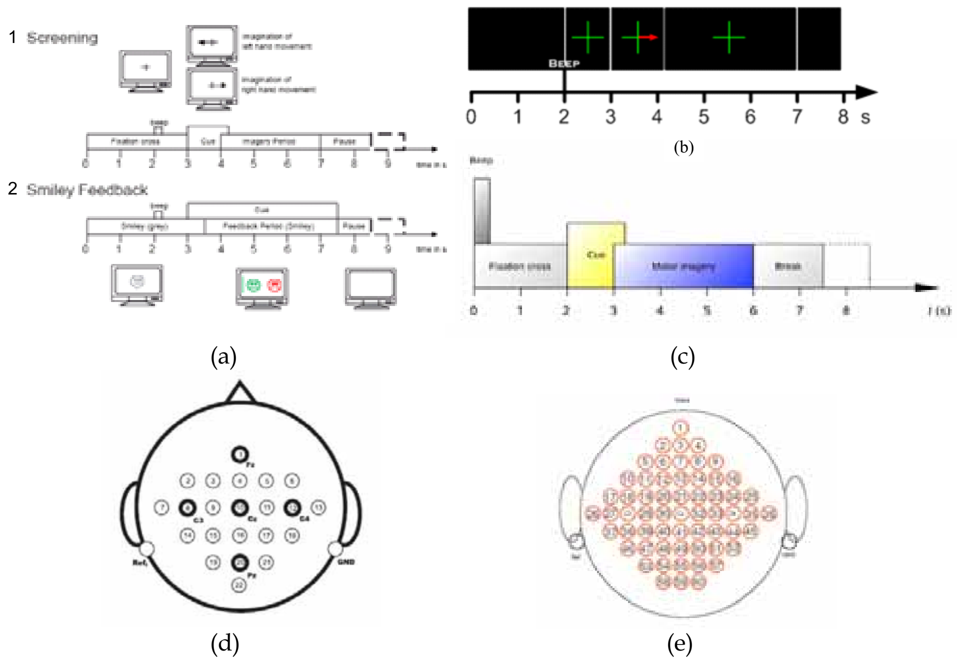
[1] Certain classifiers can work better depending on the number of dimensionality of the feature space and the number of data samples (feature vectors) available (Tebbens and Schlesinger, 2006).

analysis are presented. Section 4 contains results, including a signals and separability analysis, individual subject analysis and a statistical analysis of the methods presented. A discussion of results is presented in Section 6 which also concludes the chapter.

2. Data Acquisition and Datasets

Data from 23 subjects is used in this work. All datasets were obtained from the third and fourth international BCI competitions, BCI-III (Blankertz et al., 2005) and BCI-IV (Blankertz et al., 2008), which include datasets 2A and 2B from BCI-IV (Schlogl et al., 2008a; 2008b) and dataset IIIa from BCI-III (Schlogl et al., 2005a; 2005b). Table 1 below provides a summary of the data.

Dataset 2B - This data set consists of EEG data from 9 subjects (S1-S9). Three bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz (downsampled to 125Hz in this work). The placement of the three bipolar recordings (large or small distances, more anterior or posterior) were slightly different for each subject (for more details see (Schlogl et al., 2008b; Leeb et al., 2007). The electrode position Fz served as EEG ground. The cue-based screening paradigm (cf. Fig. 1(a).1) consisted of two classes, namely the motor imagery (MI) of the left hand (class 1) and the right hand (class2). Each subject participated in two screening sessions without feedback recorded on two different days within two weeks. Each session consisted of six runs with ten trials each and two classes of imagery. This resulted in 20 trials per run and 120 trials per session. Data of 120 repetitions of each MI class were available for each person in total. Prior to the first motor imagery training the subject executed and imagined different movements for each body part and selected the one which they could imagine best (e. g., squeezing a ball or pulling a brake). For the three online feedback sessions four runs with smiley feedback were recorded whereby each run consisted of twenty trials for each type of motor imagery (cf. Fig. 1(a).2 for details of the timing paradigm for each trial). Depending on the cue, the subjects were required to move the smiley towards the left or right side by imagining left or right hand movements, respectively. During the feedback period the smiley changed to green when moved in the correct direction, otherwise it became red. The distance of the smiley from the origin was set according to the integrated classification output over the past two seconds (more details can be found in (Leeb et al., 2007)). The classifier output was also mapped to the curvature of the mouth causing the smiley to be happy (corners of the mouth upwards) or sad (corners of the mouth downwards). The subject was instructed to keep the smiley on the correct side for as long as possible and therefore to perform the MI as long as possible. A more detailed explanation of the dataset and recording paradigm is available (Schlogl et al., 2008a). In addition to the EEG channels, the electrooculogram (EOG) was recorded with three monopolar electrodes and this additional data can be used for EOG artifact removal (Schlogl et al., 2007b) but was not used in this study.



(b)

(a)

(c)

(d)

(e)

Fig. 1. (a) Timing scheme of the paradigm for recording dataset 2B; 1) the first two sessions provided training data without feedback, and 2) the last three sessions with smiley feedback. (b) Timing scheme of the paradigm for recording dataset IIIa; (c) Timing scheme of recording for dataset 2A; (d) electrode montage for recording dataset 2A; (e) electrode montage for recording dataset IIIa with the chosen subset of 22 electrodes shown (red) and electrodes used to derive bipolar channels around c3, cz and c4. For dataset 2B electrodes positions were fine tuned around positions c3, cz and c4 for each subject0 (Leeb et al., 2007)

Competition	*Dataset*	*Subjects*	*Labels*	*Trials*	*Classes*	*Channels*
BCI-IV	2B	9	S1-S9	1140	2	3
BCI-IV	2A	9	S10-S18	576	4	22
BCI-III	IIIa	3 (+2)=5	S19-S23	240-360	4	60

Table 1. Summary of datasets used from the International BCI competitions 2003 and 2008 plus additional provided datasets.

Dataset 2A - This dataset consists of EEG data from 9 subjects (S10-S18). The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4) (only left and right hand trials are used in this investigation). Two sessions were recorded on different days for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session. The timing scheme of one trial is illustrated in Fig. 1(c). The subjects sat in a comfortable armchair in front of a computer screen. No feedback was provided but a cue arrow indicated which motor imagery to perform. The subjects were asked to carry out the motor imagery task according to the cue and timing presented in Fig. 1(c). For each subject twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG; the montage is shown in Fig. 1(d) left. All signals were recorded monopolarly with the left mastoid serving as reference and the right mastoid as ground. The signals were sampled with 250 Hz (downsampled to 125Hz in this work) and bandpass filtered between 0.5 Hz and 100 Hz. EOG channels were also recorded for the subsequent application of artifact processing although this data was not used in this work. A visual inspection of all data sets was carried out by an expert and trials containing artifacts were marked. For a full description of the recording procedure see (Schlogl et al., 2008b).

Dataset IIIa – This dataset was recorded from three subjects, S19-S21 using a 64-channel Neuroscan amplifier (datasets with the same recording procedure obtained from 2 additional subjects were provided by the organizers after the competition (S22-S23)). Sixty EEG channels were recorded using a 250Hz sampling rate (down-sampled to 125Hz in this work). The electrode positioning is illustrated in Fig. 1 (e). The training involved the sequential repetition of a cue based trial according to the paradigm and timing illustrated in Fig. 1(b) for each of the 5 subjects. The subjects were seated in a comfortable chair and instructed to imagine left hand, right hand, foot, or tongue movement according to the direction of the cue arrow on the screen (only left and right hand trials are used in this investigation). Each of the four motor imagery tasks was performed 10 times within each run in a randomized order. In this experiment no feedback was provided to the subject. Subjects 1 performed 360 and subjects 2-5 performed 240 trials (cf. Schlogl et al., 2005a; 2005b for further details).

To summarize, in this work only twenty of the sixty available channels for dataset IIIa are used as shown in Fig. 1(e). For all datasets 2 channel and 3 channel montages were also tested using the electrodes positioned anteriorly and posteriorly to c3, cz and c4 positions to derive 2-3 bipolar channels (i.e., the 2 channel montage involves c3 and c4, whereas the 3 channel montage also included cz). These channels are located over left, right hemisphere and central sensorimotor areas – areas which are predominantly the most active during motor imagery. As outlined all data was downsampled to 125 Hz in this work also.

3. Methods

3.1 Neural-Time-Series-Prediction-Preprocessing

NTSPP, introduced in (Coyle et al., 2005a), is a framework specifically developed for preprocessing EEG signals. NTSPP increases data separability by predictive mapping and filtering the original EEG signals to a higher dimensional space using predictive/regression models specialized (trained) on different EEG signals. The basic concept behind NTSPP is focused around exploiting the differences in prediction outputs produced by different predictor networks specialized on predicting different types of EEG signals to help improve the separability of EEG data and enhance overall BCI performance.

Consider two EEG times-series, x_i, $i \in \{1,2\}$ drawn from two different signal classes c_i, $i \in \{1,2\}$, respectively, assuming, in general, that the time series have different dynamics in terms of spectral content and signal amplitude but have some similarities. Consider also two prediction neural networks, f_1 and f_2, where f_1 is trained to predict the values of x_1 at time $t+\pi$ given values of x_1 up to time t (likewise, f_2 is trained on time series x_2), where π is the number of samples in the prediction horizon. If each network is sufficiently trained to specialize on its respective training data, either x_1 or x_2, using a standard error-based objective function and a standard training algorithm, then each network could be considered an ideal predictor for the data type on which it was trained[2] i.e., specialized on a particular data type.

In such cases the expected value of the mean error residual given predictor f_1 for signal x_1 is $E[x_1 - f_1(x_1)]=0$ and the expected power of the error residual, $E[x_1 - f_1(x)]^2$, would be low whereas, if x_2 is predicted by f_1 then $E[(x_2 - f_1(x_2)] \neq 0$ and $E[(x_2 - f_1(x_2)]^2$ would be high. The opposite would be observed when x_i, $i \in \{1,2\}$ data are predicted by predictor f_2. Based on the above assumptions, a simple set of rules could be used to determine which signal class an unknown signal type, u, belongs too. To classify u one, or both, of the following rules could be used

1. If $E[u - f_1(u)] = 0$ & $E[u - f_2(u)] \neq 0$ then $u \in C_1$, otherwise $u \in C_2$.

2. If $E[u - f_1(u)]^2 < E[u - f_2(u)]^2$ then $u \in C_1$, otherwise $u \in C_2$.

These rules are simple rules and may only work successfully in cases where the predictors are ideal. Due to the complexity of EEG data and its non-stationary characteristics, and the necessity to specify an NN architecture which approximates universally, predictors trained on EEG data will not consistently be ideal however; when trained on EEG with different dynamics e.g., left and right movement imagination (left or right motor imagery), predictor networks can introduce desirable characteristics in the predicted outputs which render them more separable than the original signals and thus aid in determining which class an unknown signal belongs to. As is shown in Section 3 this predictive filtering alters levels of variance in the predicted signals for data types and most importantly manipulates the variances differently for different classes. Instead of using only one signal channel, the hypothesis underlying the NTSPP framework is that if two or more channels are used for each signal class and more advanced feature extraction techniques and classifiers are used instead of the simple rules outlined above, additional useful information relevant to the differences introduced by the predictors for each class of signal (where the networks have been trained to specialise on particular data dynamics) can be extracted to improve overall feature separability and thus produce features that are easier classified than the original signals.

In general, the number of time-series available and the number of classes governs the number of predictor networks that must be trained and the resultant number of predicted time series from which to extract features,

$$P = M \times C \tag{1}$$

where P is the number of networks (=no. of predicted time-series), M is the no. of EEG channels and C the is number of classes. For prediction, the recorded EEG time-series data is structured so that the signal measurements from sample indices t to t-$(\Delta$-$1)\tau$ are used to

[2] Multilayered feedforward NNs and adaptive neuro fuzzy inference systems (ANFIS) are considered universal approximators due to having the capacity to approximate any function to any desired degree of accuracy with as few as one hidden layer that has sufficient neurons (Hornik et al., (1989); Jang et al., 1997).

make a prediction of the signal at sample index $t+\pi$. Parameter Δ is the embedding dimension and

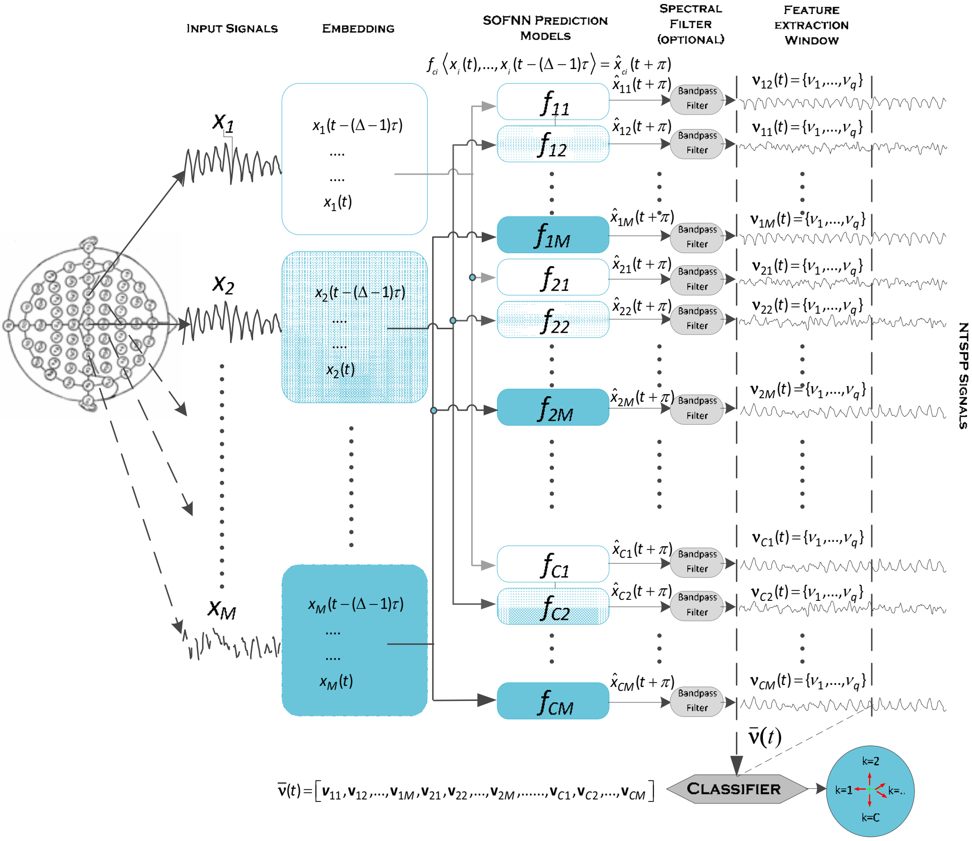


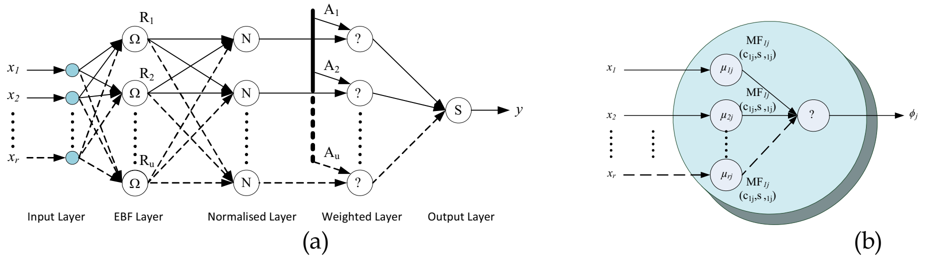
Fig. 2. Illustration of a generic multiclass or multichannel neural-time-series-prediction-preprocessing (NTSPP) framework with spectral filtering, feature extraction and classification.

$$\hat{x}_{ci}(t+\pi) = f_{ci}\langle x_i(t),...,x_i(t-(\Delta-1)\tau)\rangle \tag{2}$$

where τ is the time delay, π is the prediction horizon, f_{ci} is the prediction model trained on the i^{th} EEG channel, i=1,..,M, for class c, c=1,..C, x_i is the EEG time-series from the ith channel and $\hat{x}_{ci}$ is the predicted time series produced for the channel i by the predictor for class c, channel i. An illustration of the NTSPP framework is presented in Fig. 2.

Many different predictive approaches can be used for prediction in the NTSPP framework (Coyle, 2006). In this work the self-organizing fuzzy neural network (SOFNN) is employed (Coyle et al., 2006; 2009; Leng, 2003; Prasad et al., 2008). This is a powerful prediction algorithm capable of self-organizing its architecture, adding and pruning neurons as required. New neurons are added to cluster new data that the existing neurons are unable to cluster (cf. the following section for further details). Fine tuning parameters such as the Δ and τ may enhance the predictive performance and/or BCI performance but earlier work (Coyle et al., 2005a, Coyle 2006) has shown Δ=6 and τ=1 provide good performance in a two class motor imagery BCI and these values are used in this investigation. The SOFNNs are easily trained using a 3s window of event-related segments of signals drawn from between 1-10 randomly chosen, artifact free trials. Trials containing artifacts were not used to train the networks because artifact contaminated trials can prevent the networks from specializing on a particular motor imagery.

3.2 The Architecture of the SOFNN



(a) (b)

Fig. 3. (a) The architecture of the self-organising fuzzy neural network (b) Structure of the *j*th neuron R_j within the EBF layer

The SOFNN is a five-layer fuzzy NN and has the ability to self-organize its neurons in the learning process for implementing TS fuzzy models (Takagi and Sugeno., 1985) (cf. Fig. 3(a)). In the EBF layer, each neuron is a T-norm of Gaussian fuzzy MFs belonging to the inputs of the network. Every MF thus has a distinct centre and width, therefore every neuron has a centre and a width vector. Fig. 3(b) illustrates the internal structure of the *j*th neuron, where the input vector is $\boldsymbol{x} = [x_1\, x_2 \dots x_r]$, $\boldsymbol{c_j} = [c_{1j}\, c_{1j} \dots c_{rj}]$ is the vector of centers in the *j*th neuron, and $\boldsymbol{\sigma_j} = [\sigma_{1j}\, \sigma_{2j} \dots \sigma_{rj}]$ is the vector of widths in the *j*th neuron. Layer 1 is the input layer with *r* neurons, x_i, i=1,2,…,*r*. Layer 2 is the EBF layer. Each neuron in this layer represents a premise part of a fuzzy rule. The outputs of (EBF) neurons are computed by products of the grades of MFs. Each MF is in the form of a Gaussian function,

$$\mu_{ij} = \exp\left[-(x_i - c_{ij})^2 \big/ 2\sigma_{ij}^2\right] \qquad j = 1, 2, \cdots, u \tag{3}$$

where, μ_{ij} is the *i*th MF in the *j*th neuron;
c_{ij} is the centre of the *i*th MF in the *j*th neuron;
σ_{ij} is the width of the *i*th MF in the *j*th neuron;
r is the number of input variables;
u is the number of EBF neurons.

For the *j*th neuron, the output is

$$\phi_j = \exp\left[-\sum_{i=1}^{r}\left((x_i - c_{ij})^2 \big/ 2\sigma_{ij}^2\right)\right] \qquad j = 1,2,\cdots,u. \tag{4}$$

Layer 3 is the normalized layer. The number of neurons in this layer is equal to that of layer 2. The output of the *j*th neuron in this layer is

$$\psi_j = \phi_j \Big/ \sum_{k=1}^{u} \phi_k \qquad j = 1,2,\cdots,u. \tag{5}$$

Layer 4 is the weighted layer. Each neuron in this layer has two inputs and the product of these inputs as its output. One of the inputs is the output of the related neuron in layer 3 and the other is the weighted bias w_{2j}. For the TS model (Takagi and Sugeno., 1985), the bias $\mathbf{B}$=[1,x_1, x_2,…, x_r]^T and $\mathbf{A_j}$=[a_{j0},a_{j1},a_{j2},…,a_{jr}] represent the set of parameters corresponding to the consequent of the fuzzy rule *j* which are obtained using the least square estimator or recursive LSE (RLSE). The weighted bias w_{2j} is

$$w_{2j} = \mathbf{A_j}.\mathbf{B} = a_{j0} + a_{j1}x_1 + \cdots + a_{jr}x_r \qquad j = 1,2,\cdots,u. \tag{6}$$

This is the consequent part of the *j*th fuzzy rule of the fuzzy model. The output of each neuron is f_j= $w_{2j}\psi_j$. Layer 5 is the output layer where the incoming signals from layer 4 are summed, as shown in (7)

$$y(\mathbf{x}) = \sum_{j=1}^{u} f_j \tag{7}$$

where, *y* is the value of an output variable. If *u* neurons are generated from *n* training exemplars then the output of the network can be written as

$$\mathbf{Y} = \mathbf{W_2}\boldsymbol{\Psi}\,. \tag{8}$$

where for the TS model

$$\mathbf{Y} = [y_1 \quad y_2 \quad \cdots \quad y_n], \tag{9}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \cdots & \psi_{1n} \\ \psi_{11}x_{11} & \cdots & \psi_{1n}x_{1n} \\ \vdots & \vdots & \vdots \\ \psi_{11}x_{r1} & \cdots & \psi_{1n}x_{rn} \\ \vdots & \vdots & \vdots \\ \psi_{u1} & \cdots & \psi_{un} \\ \psi_{u1}x_{11} & \cdots & \psi_{un}x_{1n} \\ \vdots & \vdots & \vdots \\ \psi_{u1}x_{r1} & \cdots & \psi_{un}x_{rn} \end{bmatrix}, \tag{10}$$

and

$$\mathbf{W_2} = [a_{10} \quad a_{11} \cdots \quad a_{1r} \quad \cdots \quad a_{u0} \quad a_{u1} \quad \cdots \quad a_{ur}]. \tag{11}$$

$\mathbf{W}_2$ is the parameter matrix and ψ_{jt} is the output of the jth neuron in the normalized layer for the tth training exemplar.

3.3 The SOFNN Learning Algorithm

The learning process of the SOFNN includes structure learning and parameter learning. The structure learning process attempts to achieve an economical network size by dynamically modifying, adding and/or pruning neurons. There are two criteria to judge whether or not to generate a new EBF neuron – the system *error* criterion and the *if-part* criterion. The *error* criterion considers the generalization performance of the overall network. The *if-part* criterion evaluates whether existing fuzzy rules or EBF neurons can cluster the current input vector suitably. The SOFNN pruning strategy is based on the optimal brain surgeon (OBS) approach (Hassibi and Stork, 1993). Basically, the idea is to use second derivative information to find the least important neuron. If the performance of the entire network is accepted when the least important neuron is pruned, the new structure of the network is maintained.

This section provides only a basic outline of the structure learning process, the complete structure and weight learning algorithm for the SOFNN is detailed in (Leng, 2003; Prasad et al., 2008). It must be noted that the neuron modifying, adding and pruning procedures are fully dependent upon determining the network error as the structure changes therefore a significant amount of network testing is necessary – to either update the structure based on finalized neuron changes or simply to check if a temporarily deleted neuron is significant. This can be computationally demanding and therefore an alternative approach which minimizes the computational cost of error checking during the learning process is described in (Coyle et al., 2009). A comparison of the SOFNN to the well known DENFIS is outlined in (Kasobov and Song, 2002) and it is shown that the SOFNN compares favorably to other evolving fuzzy systems in terms of structural compactness and accuracy in a range of standard benchmark tests and EEG prediction. The advantage of using the SOFNN in a BCI involving the NTSPP framework is that it has a self organizing structure and can therefore adapt autonomously to each of the time series for each class and for each subject without any parameter tuning. There are 5 standard predefined parameters of the SOFNN which govern the accuracy and complexity. The investigation presented in (Coyle et al., 2009) shows that parameters chosen via a sensitivity analysis generalize well for all subjects and all signals and these parameter values have been used in this work to apply the SOFNN autonomously.

3.4 Common Spatial Patterns(CSPs)

The CSP method, first applied for detection of abnormalities (Ramouser et al., 2000) has been used to tackle the problem of extracting the most relevant information from multiple electrode (multichannel) montages. The goal of the study in (Ramouser et al., 2000) was to design spatial filters that produce new (surrogate) time-series of which the variances are optimal for the discrimination of two classes of EEG related to left and right motor imagery.

Many advances in the CSP methods have been proposed over the past few years and this approach has shown significant potential for two-class BCIs ((Blankertz et al., 2008; Coyle et al., 2008a; Dornhege et al., 2006; Ramouser et al., 2000; Satti at al., 2008; 2009).

To utilise CSP, let Σ_1 and Σ_2 be the pooled estimates of the covariance matrices for two classes, as follows:

$$\Sigma_c = \tfrac{1}{I_c}\sum\nolimits_{i=1}^{I_c} X_i X_i^t \quad (c \in \{1,2\}) \tag{12}$$

where I_c is the number of trials for class c and X_i is the M×N matrices containing the i^{th} windowed segment of trial I; V is the window length and M is the number EEG channels – when CSP is used in conjunction with NTSPP, M=P as per (1). The two covariance matrices, Σ_1 and Σ_2, are simultaneously diagonalized such that the eigenvalues sum to 1. This is achieved by calculating the generalised eigenvectors W:

$$\Sigma_1 W = (\Sigma_1 + \Sigma_2)WD \tag{13}$$

where the diagonal matrix D contains the eigenvalues of Σ_1 and the column vectors of W are the filters for the CSP projections (Blankertz et al., 2008). With this projection matrix the decomposition mapping of the windowed trials X is given as

$$E = WX \tag{14}$$

Prior to the calculation of the spatial filters, X can be processed with NTSPP and/or spectrally filtered in specific frequency bands. Many studies have shown that subject-specific frequency bands are most appropriate (Blankertz et al., 2008; Pfurtscheller et al., 1998; Pfurtscheller, 1998; Coyle et al, 2005b; Herman et al., 2008) and are normally tuned by heuristic search with a 1 Hz resolution however; in this work, to minimize the effort and time required in performing an extensive search for the best subject-specific frequency bands, only 4 bands between 8-24Hz were tested (i.e., 8-12; 8-16; 8-20, 8-24). These bands encompass the μ and β bands which are altered during sensorimotor processing (Pfurtscheller et al., 1998; Pfurtscheller, 1998). Attenuation of the spectral power in these bands indicates an event related desynchronization (ERD) whilst an increase in power indicates event-related synchronization (ERS). ERD of the mu band or ERS of the beta band is associated with activated sensorimotor areas and ERS in the mu band is associated with idle or resting sensorimotor areas. ERD/ERS has been studied widely for many cognitive studies and provides very distinctive lateralized EEG pattern differences which form the basis of left/right motor imagery based BCIs (Pfurtscheller, 1998).

3.5 Feature Extraction

Features are extracted using a 1 second window through which the data for each trial is passed either via NTSPP or the raw EEG signals and classified at rate of the sampling interval. These signals X are decomposed according to (14) and each feature vector, $\overline{v}$, is obtained using (15).

$$\overline{v} = \log(\mathrm{var}(E)) \tag{15}$$

The dimensionality of $\overline{v}$ depends on the number of surrogate signals used from E. The common practice is to use several (between 2 and 6) eigenvectors from both ends of the eigenvector spectrum, i.e., the columns of W. As can be seen from Fig. 2, if NTSPP is performed the dimensionality of X can increase as shown in (1) and becomes N×P. Depending on the number of classes and the number of signals available, the dimensionality increase can be significant. NTSPP maps the original data to a higher dimensional signal space which is more separable but also susceptible to containing redundant information in addition to increasing the dimensionality of the feature vector after features are extracted from the NTSPP (i.e., predicted) signals. Large feature vectors can result in sparse matrices for training certain classifiers when the number of exemplars is low. This can significantly impact on the performance of certain classifiers (Tebbens & Schlesinger, 2006). CSP on the other hand can be used to reduce the dimensionality of the available data and also perform a further mapping of the data to increase separability. Therefore the benefits of combining NTSPP with CSP are two fold:- 1) increasing separability and 2) maintaining a tractable dimensionality.

To quantify these benefits and the benefits of employing CSP in BCI with a low number of channels, which is not normally done in BCI, the following tests have been carried using a 2 channel montage, a 3 channel montage and a 22 channel montage as shown in Fig. 1.

- SF – spectral filtering only as a benchmark (2 and 3 channel montages only)
- SF-CSP – spectral filtering and common spatial patterns which is a normal BCI setup
- NTSPP-SF – NTSPP and spectral filtering to show the performance of NTSPP compared to CSP as a standalone preprocessing tool (2 and 3 channel montages only)
- NTSPP-SF-CSP – a combination of all preprocessing methods

Tests are not performed for the SF and NTSPP-SF tests using a 22 channel montage because without CSP the dimensionality of the feature vectors is 22 for SF (22 channels) and 44 for NTSPP-SF (22 channels x 2 classes as shown in (1)). As outlined, without employing CSP, the dimensionality of such feature vectors and the redundancy and/or noise in some channels could impact on the overall performance and therefore some method of feature selection/channel reduction is necessary. When CSP is employed tests are carried out using up to a maximum of 4 eigenvectors from either end of W. Depending on the number of EEG channels available and whether or not NTSPP is employed there are different amounts of eigenvectors to choose from and choosing the optimum number can often impact on performance therefore; when the option to have less or more eigenvectors was available, tests were performed with each number. For example, when a 2 channel montage is employed the maximum number of available eigenvectors is 1 from either end of W for SF-CSP and 2 for NTSPP-SF-CSP therefore tests are performed once with SF-CSP and 2 times with NTSPP-SF-CSP and so on.

3.6 Classification

Four different classifiers obtained from the Biosig toolbox (Schlogl, 2009) are used with all methods described. These include linear discriminant analysis (LDA), support vectors machines (SVMs), Mahalanobis distance classifier (MDA) and a generalized distance based classifier (GDBC) (cf. (Schlogl, 2009) for further details). In addition, a probabilistic Bayes based classifier involving the accumulation of evidence was employed (cf. (Duda et al., 2001;

Lemm et al., 2004) for further details). By using each of these classifiers a better general view of each methods performance was attained.

The datasets for each subject were split into two sets where half the data is used for training and validation and the other half used for testing. These tests are referred to as *5-fold* and *single trial test* sets. Using each of the 6 classification methods, a 5-fold cross-validation was carried out on the 5-fold set for each subject, where the data was partitioned into a training set (80%) and a validation set (20%). Tests were performed five times using a different validation partition each time. The mean-CA (mCA) rates on the 5-folds of validation data and 95% confidence intervals (ci) were estimated using a t-statistic. The purpose of the 5-fold cross validations was to tune any parameters and identify the point at which each subject maximized the separability between the two classes. Subsequently, all 5-fold data was utilized to train the system and the classifier was set up on the features which produced the highest mCA rate in the cross-validation on SP1. The system's generalization abilities were then tested on a one-pass single trial test on the test set – this final test corresponds to the requirement of labeling the data in online single trials test for a practically useful BCI system.

4. Results

4.1 Signals and separability analysis

To illustrate how each method enhances separability in the data for each subject a range of separability measures and visualization methods were applied to the data of each subject. Using the mean CA (mCA) on the 5-fold train and validation sets to identify the point of maximum separability, features were extracted at this time point using signals preprocessed by each of the methods from all available data (this analysis was carried out after BCI tests were performed). Using the features extracted from each signal[3] boxplots were estimated to attain a quick impression of the features' variability within and across classes, as shown in Fig. 4.

As can be seen from Fig. 4 there is substantially more interclass variability when NTSPP is employed and the NTSPP process does result in producing different median values for each of six features. The scales are different when CSP is employed so if the medians of the features obtained using the SF-CSP methods are compared with those obtained using NTSPP-SF-CSP, it can be observed that NTSPP has changed the median values of the features (i.e., features are derived using the variance calculation) and it is clear that there is more opportunity to enhance interclass variability when using NTSPP as opposed to no NTSPP. Notches display the variability of the median between samples. The width of a notch is computed so that box plots whose notches do not overlap have different medians at the 5% significance level. The significance level is based on a normal distribution assumption. Comparing box plot medians is like a visual hypothesis test, analogous to the t-test used for means and therefore it can be seen that the differences in the features produced by different NTSPP signals are significant in many cases (MATLAB®, 2009).

To quantify the separability enhancement for this subject a range of separability indices were estimated (as shown in Table 2), including the Euclidean distance (edist) between class

[3] Signals are c3, cz and c4 when no NTSPP is employed or signals are prefixed by the first letter of the data class that each predictor is trained on when NTSPP is employed i.e., l3, l4, and lz for the data processed by the left predictors and r3, r4, and rz for data processed by the right predictors.

means for which the objective is to maximize, the Davies-Bouldin index (dbi) which is a cluster separability index (Davies and Bouldin, 1979) for which the objective is to minimize, dtc is a statistical measure of the multivariate distance of each observation (feature vector).

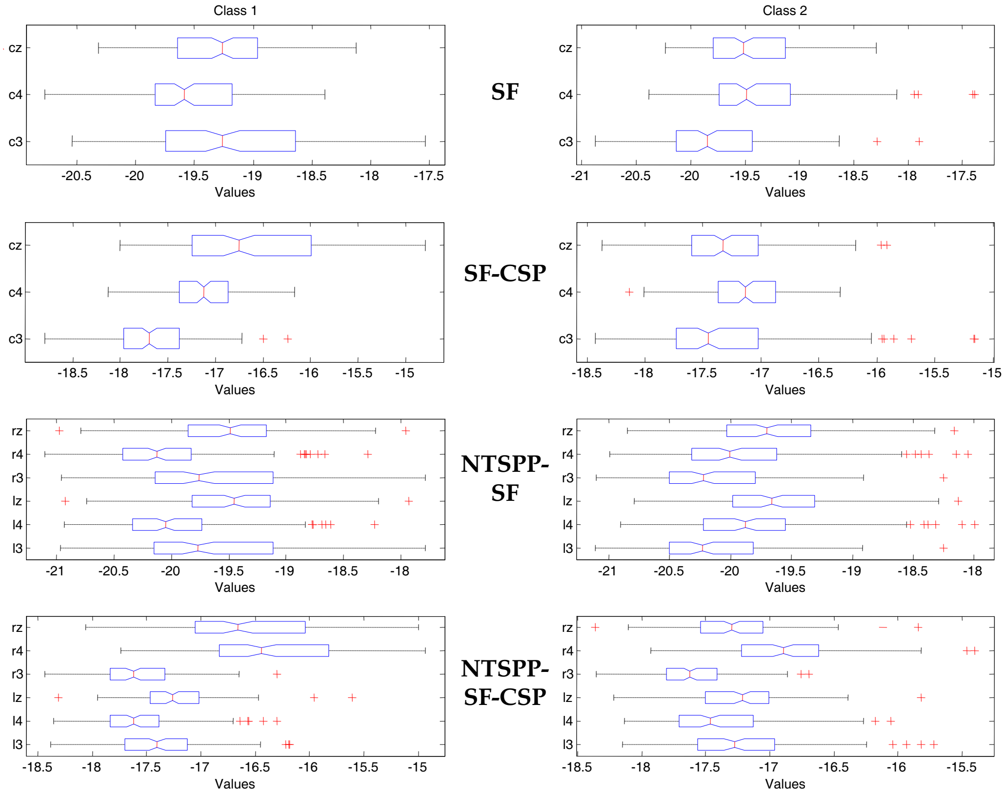


Fig. 4. Boxplots of the features extracted from each signal, for each class and for each methodology from the center of the dataset (both classes) and the class separability index (csi) is a measure of the average distance between each observation within class 1 to the centre of class 2 and vice versa.

			NTSPP-	NTSPP-
	SF	**SF-CSP**	**SF**	**SF-CSP**
mCA	76.43	76.43	78.57	80.00
edist	0.67	0.75	0.86	0.96
dbi	33.25	28.99	38.57	27.44
dtc	3.27	3.54	4.94	3.71
csi	1.87	2.09	2.19	2.10

Table 2. A range of separability indices for 1 subject for each of the methods (details of separability indices are presented in the text).

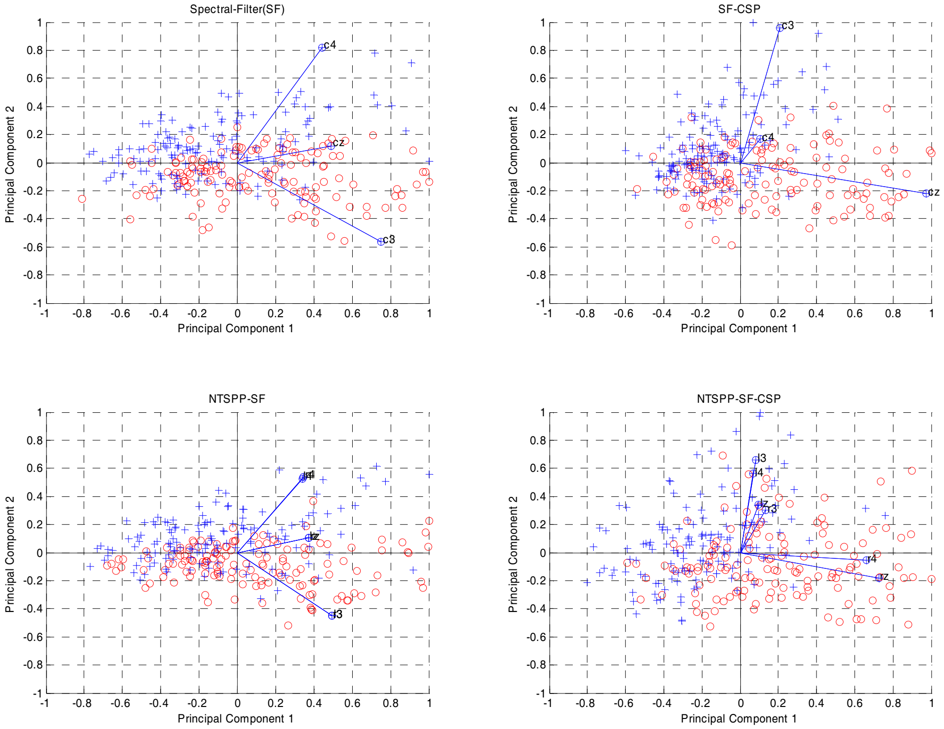


Fig. 5. Biplots showing the first 2 principle components for each of the 4 methods for 1 subject.

From Table 2 it can be seen that NTSPP produces the highest mCA on the 5 fold cross-validation. NTSPP also produces the highest separability across the data in terms of maximizing edist, minimizing dbi, and maximizing dtc and csi. It can be seen that SF alone is the worst performer on all tests, whilst SF-CSP performs better than NTSPP-SF only in dbi. Maximization of Euclidean distance with NTSPP-SF-CSP appears to be a significant benefit of employing this combination of processes which is reflected in the mCA rate which is ~4% greater than the mCA for SF-CSP with no NTSPP for this subject. With no CSP employed, NTSPP is shown to be a better preprocessor than CSP for this subject with the NTSPP-SF approach achieving higher separability than both approaches without NTSPP. The significance of the mCA results across all subjects is shown in the following section.
To aid in visualizing the multidimensional data a principle component analysis (PCA) was carried out. The two most important components for classification are shown in Fig. 5 where biplots showing the first two principle component coefficients are presented. The biplots helps visualize both the principal component coefficients for each variable and the principal component scores for each observation in a single plot.

Each of the features extracted from each signal for each method are represented in these plots by a vector, and the direction and length of the vector indicates how each variable contributes to the two principal components in the plot. The first principal component in each biplot is represented by the horizontal axis and has positive coefficients for all features for each method corresponding to the 3(6 for NTSPP) vectors directed into the right half of the plot. The second principal component, represented by the vertical axis, has positive coefficients for features obtained from c4 and cz for SF, c3 and c4 for SF-CSP, r4, l4, rz, lz for NTSPP-SF and l3, l4, lz and r3 for NTSPP-SF-CSP and has negative coefficients for the remaining five variables. This corresponds to vectors directed into the top and bottom halves of the plot, respectively. This indicates that this component distinguishes between classes that produce high values for the first set of features and low for the second, and classes that have the opposite. Overall it can be seen that the NTSPP-SF-CSP has at least 3 features which are distinguishably providing high variance for one class and two features which are providing lower variance for the other class whereas the other methods have less features that are providing this overall difference in variability, which is providing the superior separability given by NTSPP-SF-CSP in this example. This section has provided a general overview of the dynamical changes which are introduced by these NTSPP methods and the advantages produced in terms of improved separability. The following sections provide further verification of these results by providing a qualitative and statistical analysis of each of the methods when applied across the data from 23 subjects.

4.2 Classification accuracy analysis

4.2.1 Individual subject results

As per the data description in section 2 and section 3.6, results for 5-fold cross validation were obtained for all subjects. Parameter information and time point of maximum separability obtained from the cross validation were used to set up the methods for tests on the test set (single trial test), results of which provide a good indicator for online BCI performance. As outlined the objectives of the research was to compare all methods when employed with 2, 3 or 22 channels. Results for all subjects and all methods are presented in Fig. 6-Fig. 13. Multichannel datasets were not available for subjects S1-S9 therefore only results for 2 channel and 3 channel montages are presented in Fig. 6-Fig. 9. Results for subjects S10-S23 are compared for the 22 channel montages also and these results are presented Fig. 10-Fig. 13. The 22 channel results in Fig. 10 and Fig. 11 are reproduced in Fig. 12 and Fig. 13 for ease of comparison with either the 2 channel or 3 channel results respectively. Results for the Bayes based classifier and the LDA classifier provided the maximum performance in the majority of cases in the cross validation tests therefore only results for these classifiers are presented however support vectors machines (SVMs), Mahalanobis distance classifier (MDA) and a generalized distance based classifier (GDBC) did provide similar results for certain subjects (the following section provide further information on classifier performances). For SVM the regularization parameter was not tuned.

It can be seen from the results that there is quite a lot variation across subjects but in the majority of cases the accuracies for NTSPP approaches are higher than the accuracies obtained when no NTSSP is involved. The differences in accuracies are more prominent for some subjects than others and in a small number of cases the NTSPP produces lower accuracies. A statistical analysis is provided in the following section to verify the significance of the differences among each of the methods. There is a particularly noticeable increase in accuracy for the majority of subjects when 22 channels are used indicating that a 22 channel montage is much better than a 2 or 3 channel montage, however, in a number of cases 3 channel NTSPP methods produce better than or comparable performances to the 22 channel montages and in almost all cases, reduce the difference between the 3 channel results and the 22 channel results substantially more than when no NTSPP is performed using the three channel montages. These results are certainly indicative that NTSPP can improve the performance when a low number of channels are used. Again, the significance of these results is analyzed in the following section.

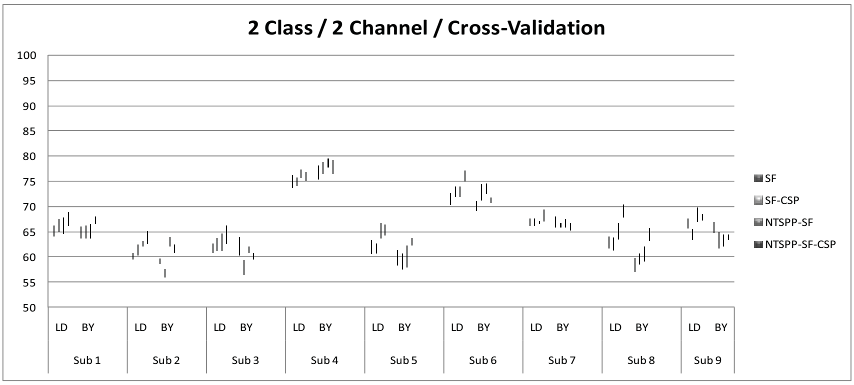


Fig. 6. mCA[%] obtained from cross validation with error bars showing the 95% confidence interval (subjects S1-S9, 2 channel).

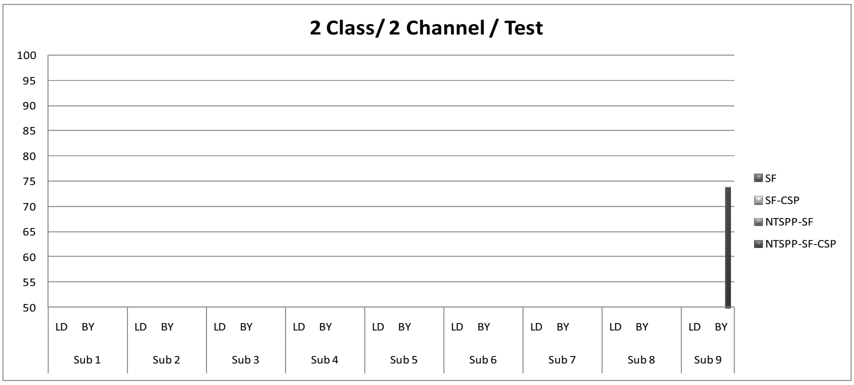


Fig. 7. CA[%] obtained from single trial tests (subjects S1-S9, 2 channel)

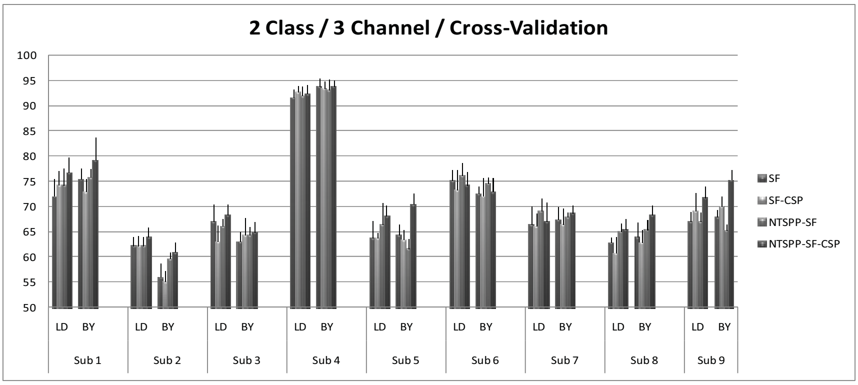


Fig. 8. mCA[%] obtained from cross validation with error bars showing the 95% confidence interval (subjects S1-S9, 3 channel)

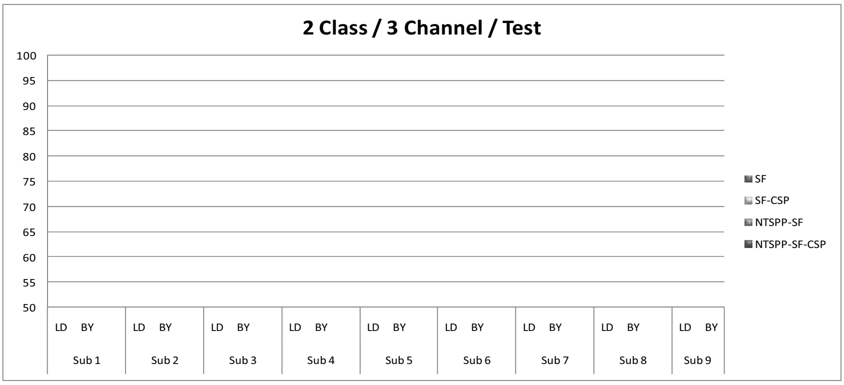


Fig. 9. CA[%] obtained from single trial tests (subjects S1-S9, 3 channel)

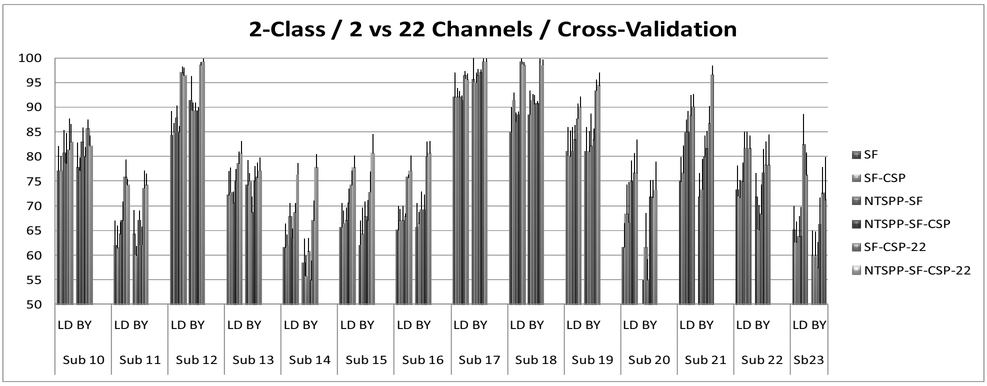


Fig. 10. mCA[%] obtained from cross validation with error bars showing the 95% confidence interval (subjects S10-S23, 2 versus 22 channel results shown)

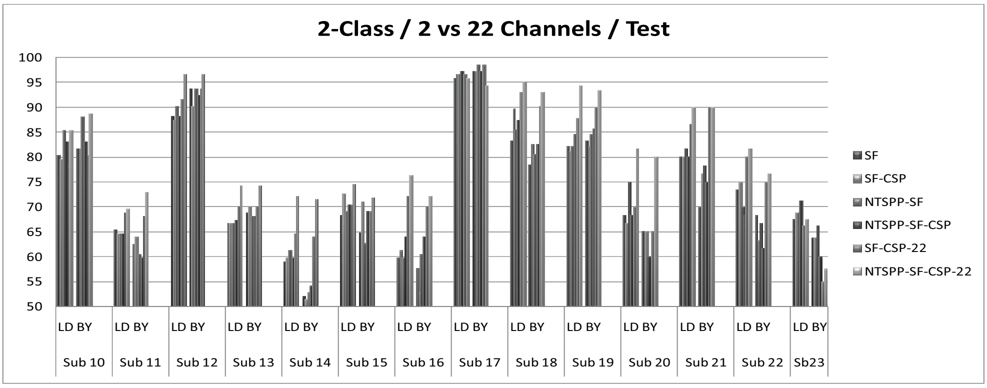


Fig. 11. CA[%] obtained from single trial tests (subjects S10-S23, 2 versus 22 channel results)

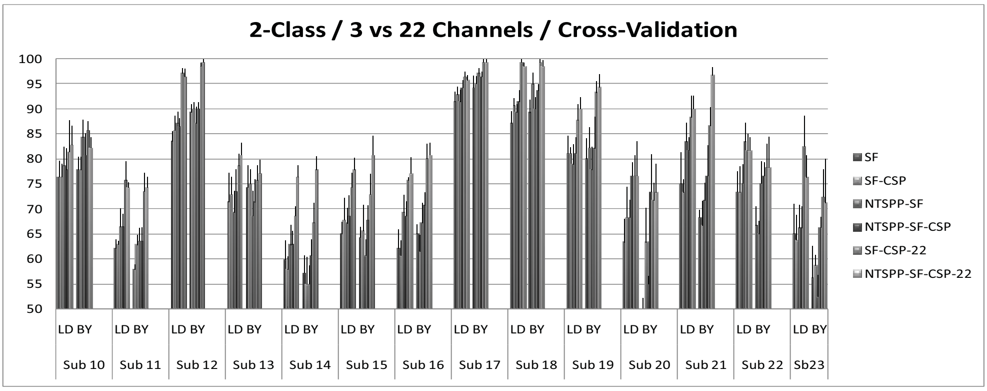


Fig. 12. mCA[%] obtained from cross validation with error bars showing the 95% confidence interval (subjects S10-S23, 3 channel versus 22 channel results shown)

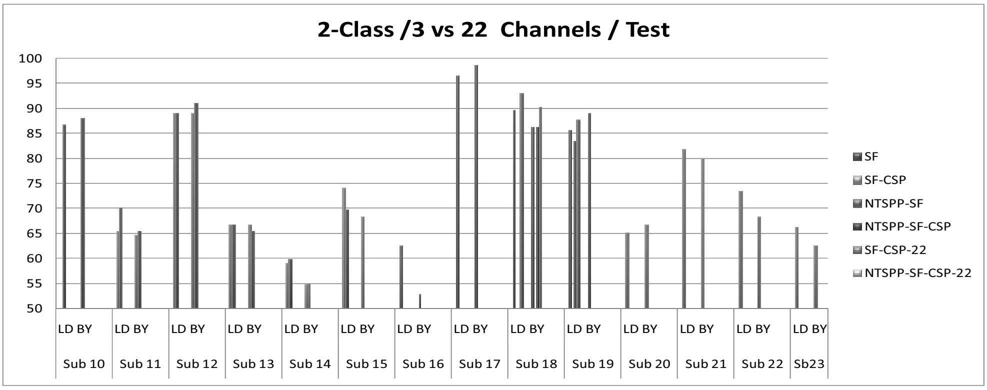


Fig. 13. CA[%] obtained from single trial tests (subjects S10-S23, 3 versus 22 channel results)

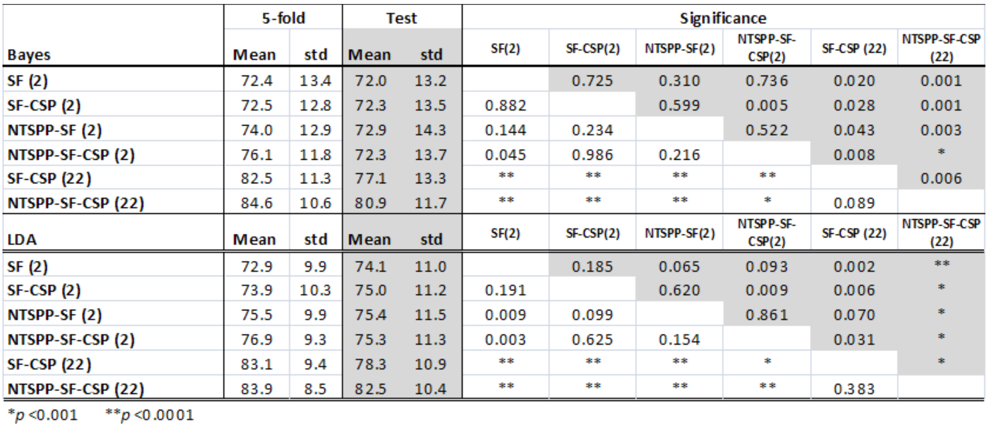
4.2.2 Statistical analysis

The results for each subject presented in the previous section show trends that NTSPP can produce better performances in many cases however there is a need to analyze all results in terms of their statistical significance, to verify whether one method is better than the other. To do this, the average accuracies for all methods across all subjects were subjected to repeated measures single factor analysis of variance (RANOVA) (Zar, 1999; Huck, 2000). This repeated measures method was preferred over standard ANOVA to account for the between subject variability which is normally substantive in BCI experiments. In this work the objective was to determine how each method compares with each other method therefore only pair-wise comparisons of means were performed which is equivalent to multiple *t*-tests. For a more powerful analysis RANOVA could be applied to all methods and a post hoc analysis of the ANOVA results could be performed. In this analysis it is off interest if there exists differences between one method and any of the other methods with a significance level α=0.05 however, to account for the multiple comparisons, the significance level, α, must be corrected. Based on a Bonferroni correction the corrected $\alpha = \alpha\ /(k.(k\text{-}1)/2)$, where k is the number of methods to be compared (i.e., k=6) therefore p<0.003 to be significant.

Table 3 and Table 4 shows the results obtained for subjects S10-S23. Only these subjects are compared as multichannel data was unavailable for Subjects S1-S9. As can be seen, average accuracies for 22 channel montages are significantly higher than those produced by the either of the 2 or 3 channel montages (p<0.003 in all cases and in some case p<0.0001).

This is evidence that there is a significant advantage in applying more channels for this two class classification problem. Although NTSPP-SF-CSP(22) is not shown to be significantly better that SF-CSP(22) for the multichannel cross-validation data, the NTSPP-SF-CSP(22) combination is significantly better than SF-CSP(22) for the single trial tests using LDA (p<0.0001) but not for Bayes. This is a strong indication that NTSPP combined with spectral filtering and CSP generalize much better to unseen data and is better for cross session single trial tests with multiple channel montages. For the 2 and 3 channel montage the results are less consistent.

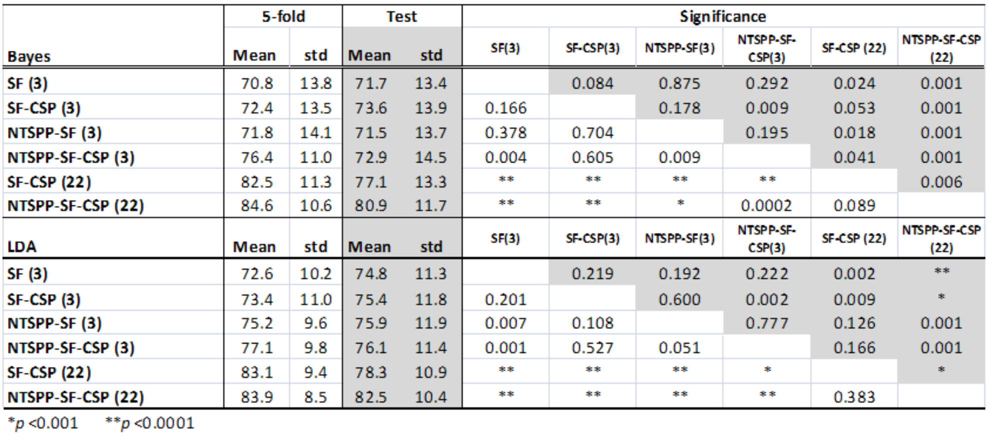
For the 2 channel montage, even though NTSPP-SF-CSP produces a higher average accuracy it is not significantly better than SF-CSP for the 5-fold data and there is only a marginal difference in performance for the single trial tests using LDA. NTSPP-SF-CSP(2) has higher mean accuracy than SF alone for cross-validation tests using the LDA classifier but the results for the single trial tests have only marginal differences. There is indication from the trends in these results that NTSPP can improve performance with 2 channel systems and in many cases the difference between NTSPP methods are significantly better than the SF methods whilst the SF-CSP methods are not significantly better than SF methods. It can also be observed from Table 3 that using a 22 channel montage the difference between SF-CSP (22) and NTSPP-SF(2) or NTSPP-SF-CSP(2) is not significant using the LDA classifier on the single trial tests whereas NTSPP-SF-CSP (22 channel) produces significant differences between all the 2 channel results using LDA and the Bayes classifiers (p<0.003 in all cases). These results indicate that the 2 channel system when employed with NTSPP-SF-CSP or NTSPP-SF and LDA can produce performances which are comparable with a 22 channel system, at least in single trial tests although the 5 fold results do not show the same trends in significance levels. Overall, even though NTSPP-SF-CSP (22 channel) produce the best results, the results do confirm that NTSPP has the potential to provide better results than SF or SF-CSP using a smaller montage also.



	5-fold		Test		Significance					
Bayes	**Mean**	**std**	**Mean**	**std**	SF(2)	SF-CSP(2)	NTSPP-SF(2)	NTSPP-SF-CSP(2)	SF-CSP (22)	NTSPP-SF-CSP (22)
SF (2)	72.4	13.4	72.0	13.2		0.725	0.310	0.736	0.020	0.001
SF-CSP (2)	72.5	12.8	72.3	13.5	0.882		0.599	0.005	0.028	0.001
NTSPP-SF (2)	74.0	12.9	72.9	14.3	0.144	0.234		0.522	0.043	0.003
NTSPP-SF-CSP (2)	76.1	11.8	72.3	13.7	0.045	0.986	0.216		0.008	*
SF-CSP (22)	82.5	11.3	77.1	13.3	**	**	**	**		0.006
NTSPP-SF-CSP (22)	84.6	10.6	80.9	11.7	**	**	**	*	0.089	
LDA	**Mean**	**std**	**Mean**	**std**	SF(2)	SF-CSP(2)	NTSPP-SF(2)	NTSPP-SF-CSP(2)	SF-CSP (22)	NTSPP-SF-CSP (22)
SF (2)	72.9	9.9	74.1	11.0		0.185	0.065	0.093	0.002	**
SF-CSP (2)	73.9	10.3	75.0	11.2	0.191		0.620	0.009	0.006	*
NTSPP-SF (2)	75.5	9.9	75.4	11.5	0.009	0.099		0.861	0.070	*
NTSPP-SF-CSP (2)	76.9	9.3	75.3	11.3	0.003	0.625	0.154		0.031	*
SF-CSP (22)	83.1	9.4	78.3	10.9	**	**	**	*		*
NTSPP-SF-CSP (22)	83.9	8.5	82.5	10.4	**	**	**	**	0.383	

*p <0.001 **p <0.0001

Table 3. Results showing the average CA rates and the standard deviation across subjects S10-S23 for the cross validation (white columns) and single trial tests (grey columns) for 2 channels and 22 channels. The significance of the differences between one method and each other method is shown in white for 5-fold cross validation and in grey for single trial tests. Only results for Bayes and LDA classifiers are presented. The significance of the difference in mean for multichannel data is also presented.



	5-fold		Test		Significance					
Bayes	**Mean**	**std**	**Mean**	**std**	SF(3)	SF-CSP(3)	NTSPP-SF(3)	NTSPP-SF-CSP(3)	SF-CSP (22)	NTSPP-SF-CSP (22)
SF (3)	70.8	13.8	71.7	13.4		0.084	0.875	0.292	0.024	0.001
SF-CSP (3)	72.4	13.5	73.6	13.9	0.166		0.178	0.009	0.053	0.001
NTSPP-SF (3)	71.8	14.1	71.5	13.7	0.378	0.704		0.195	0.018	0.001
NTSPP-SF-CSP (3)	76.4	11.0	72.9	14.5	0.004	0.605	0.009		0.041	0.001
SF-CSP (22)	82.5	11.3	77.1	13.3	**	**	**	**		0.006
NTSPP-SF-CSP (22)	84.6	10.6	80.9	11.7	**	**	*	0.0002	0.089	
LDA	**Mean**	**std**	**Mean**	**std**	SF(3)	SF-CSP(3)	NTSPP-SF(3)	NTSPP-SF-CSP(3)	SF-CSP (22)	NTSPP-SF-CSP (22)
SF (3)	72.6	10.2	74.8	11.3		0.219	0.192	0.222	0.002	**
SF-CSP (3)	73.4	11.0	75.4	11.8	0.201		0.600	0.002	0.009	*
NTSPP-SF (3)	75.2	9.6	75.9	11.9	0.007	0.108		0.777	0.126	0.001
NTSPP-SF-CSP (3)	77.1	9.8	76.1	11.4	0.001	0.527	0.051		0.166	0.001
SF-CSP (22)	83.1	9.4	78.3	10.9	**	**	**	*		*
NTSPP-SF-CSP (22)	83.9	8.5	82.5	10.4	**	**	**	**	0.383	

*p <0.001 **p <0.0001

Table 4. Results showing the average CA rates and the standard deviation across subjects S10-S23 for the cross validation (white columns) and single trial tests (grey columns) for 3 channels and 22 channels. The significance of the differences between one method and each other method is shown in white for 5-fold cross validation and in grey for single trial tests. Only results for Bayes and LDA classifiers are presented. The significance of the difference in mean for multichannel data is also presented.

For the 3 channel results presented in Table 4 it can be seen that accuracies obtained using NTSPP-SF-CSP and SF-CSP are better than those produced when CSP is not employed when using the Bayes classifier in the cross validation and single trial tests. NTSPP-SF-CSP is significantly better than SF alone for cross validation but not for the single trial tests and SF-CSP is marginally better than NTSPP-SF-CSP in single trial tests using the Bayes classifier. Using the LDA classifier NTSPP-SF-CSP is marginally better than SF-CSP but not significantly better than SF alone for the single trial tests whereas NTSPP approaches are significantly better than SF alone for the cross-validation test but not better than SF-CSP. Again for the 22 channel montages, SF-CSP(22) is not significantly better than NTSPP methods using the LDA classifier but is significantly better than SF and SF-CSP (2 channels) which indicates the potential for NTSPP to produce better results than other methods on smaller montages. The NTSPP-SF-CSP(22) methods produce results which are statistically better than all 3 channel methods which indicates that NTSPP can also enhance results even with multichannel systems.

In summary, with two and three channels some results indicate that NTSPP methods can produce similar single trial performances to the 22 channel results obtained using SF-CSP, a result which indicates that NTSPP can be used to enhance the performance of BCIs with a minimal number of electrodes, reducing the burden of mounting a multiple electrodes. The results also clearly indicate that NTSPP-SF-CSP with the 22 channel montage produces significantly better single trial results than all other methods (including SF-CSP with 22 channels) for both classifiers which are considerable evidence of the NTSPP framework's capacity to stabilize cross session tests in multiple channel systems also. When all 14 subjects are taken into consideration there is substantive evidence to suggest that NTSPP significantly enhances performances when employed with SF-CSP and in many cases also when only the NTSPP-SF combination is employed. This is indicative that NTSPP can be used instead of CSP as a preprocessing methodology but, also, that combining NTSPP and CSP in addition to spectral filtering, can lead to significant performance enhancements, regardless of the number of channels or the type of classifier used therefore NTSPP and CSP are complementary approaches. It must be noted that the Bonferroni correction is conservative correction measure for significance tests. This factor, in addition to the relatively small sample size and substantive inter subjects performance variability, can have a significant impact on measuring the statistical significance of results however the results presented do prove the significance of employing NTSPP.

In term of the classifiers, in general, the Bayes classifier overall does not produce accuracies that are as high as the LDA classifier and is less stable and this may explain why SF-CSP produced marginally better single trial results than NTSPP-SF-CSP using the Bayes classifier in a small number of cases. Although the Bayes classifier may not generalize as well as other classifiers, with accumulation of evidence overtime within each trial the Bayes approach offers better within trial stability. This is achieved by using information about the classifier output from previous time points in the trial when classifying the current time point. In the majority of cases all other classifiers provide slightly lower performance than LDA. A range of RANOVA tests were carried out and it was observed that LDA outperformed all other methods in the single trial tests and that the differences in the performances were statistically significant (p<0.05). Different overall averages were obtained depending on the data type being classified however the results do indicate that LDA is most stable for single trial tests, although both SVM and Bayes could have been improved further by fine tuning a number of regularization parameters for each subject. In this work parameter tuning was kept to a minimum and LDA has the advantage of producing the best performance with no effort required for parameter tuning. LDA is the state-of the-art for classification in two class BCI systems and these results provide further evidence of that.

5. Discussion and Conclusions

NTSPP can act as a filter of irregular transients and noise sources, since filtering and prediction go hand in hand. However NTSPP is different to basic filtering in that different filters/predictors are developed for different data types but used to process both data types. This work has shown the value of employing NTSPP as an alternative preprocessing method to the well known CSP filtering approach. CSP has been employed in BCI systems for over ten years and is employed in a range of state-of-the-art BCI systems (Blankertz et al., 2008; Dornhege et al., 2006; Ramouser et al., 2000). It has also been shown that application of NTSPP in combination with CSP has significantly more potential than either approach employed individually. For example, as outlined, when the amount of available channels is large, CSP can be used not only to produce surrogate data which maximizes the variances for one class whilst minimizing for the other class, it can act as a signal/feature selector to reduce data dimensionality. NTSPP on the other hand also manipulates the variances of the data by predictive filtering but results in a dimensionality increase, which can be significant if the number of available EEG channels and/or classes is large. The can in some cases lead to redundancy which may have implications for classifier performance if the number of available training samples is low. By applying both approaches the manipulation of variances are complementary, in addition to CSP deriving a subset of new channels from the signals predicted by NTSPP to reduce dimensionality. The results have demonstrated the advantages in doing this for both small and multichannel montages. In (Coyle et al., 2008a) NTSPP was employed with simple features in a 4 class BCI where CSP was not employed and it was noted that there was redundancy and significant dimensionality increases and thus the results were not so consistent. An analysis is underway to show the benefits of the NTSPP-CSP combination when employed in a 4 class BCI, an approach that was employed for the multiple channel dataset in the recent International BCI competition, results of which are available online (Blankertz et al., 2008b). NTSPP has also been shown to have the capacity to reduce the latency involved in motor imagery BCIs involving continuous classification; producing higher signal separability faster (i.e., earlier in the trial) by predicting the EEG times series multiple steps ahead and subsequently features are extracted from the predicted signals. This has the potential to reduce the time required for a subject to exceed a threshold with the continuous classifier output, as NTSPP predicts characteristics of the data which are more separable multiple steps ahead in time (Coyle et al., 2004, 2009) and further work will be carried out to verify if combining CSP with the multiple step ahead prediction NTSPP framework has significant potential. In terms of improving the NTSPP framework, there is a lot that can be done. For example, a more intuitive process for selecting the embedding dimension and time lag may produce predictors which are better or more specialized and thus result in producing better variability in the outputs for different classes. However simplicity is favored over complexity in BCI development, to enable easier adaptation to each individual and continuous adaptation in the long term (Wolpaw, 2004) so the number of signals and subject specific parameters should be kept to a minimum. NTSPP increases the potentiality of using simpler feature extraction methods or reducing the necessity to fine tune parameters in more complex feature extraction methods. Also, the improved autonomy in adaptation and performance offered by the self-organizing fuzzy neural network (SOFNN) allows the NTSPP framework to be applied autonomously (no parameter tuning is necessary) (Coyle et al., 2006b; 2009).

In terms of improving all methods, the spectral filters could be tuned more precisely. In this work 4 bands were tested with a wideband 8-24 Hz being most useful in some cases whilst a narrow band (8-12Hz) being better in other cases. Fine tuning of the frequency filters in concert with the preprocessing methods, as described in (Satti et al., 2009), would undoubtedly result in better performance for some subjects if not all. Nevertheless a major objective of this work is to keep to a minimum the number of subject specific parameters and the amount of time and expert knowledge required to set up the BCI system. It is unclear whether spectral filtering prior to network training would provide better results and this will also be a topic of further investigation.

Overall this work has shown the advantages and performance gain that can be produced using NTSPP as an easily applied method for preprocessing and that NTSPP, in combination with spectral filtering and common spatial patterns, can offer superior performance than any of the approaches used independently. There is lot of potential to enhance the NTSPP framework and this is part of ongoing investigations.

6. Acknowledgment

The author would like to acknowledge and thank the organizers of the BCI Competitions III and IV, Benjamin Blankertz (Blankertz et al., 2005; 2008a), and also Gert Pfurtscheller and Alois Schlogl for providing the competition datasets IIIa, 2A and 2B and additional EEG data (Schlogl et al., 2005a; 2005b; 2008a; 2008b) and the Biosig toolbox (Schlogl et al., 2009).

7. References

Birbaumer, N.; Ghanayim, N.; Hinterberger, T.; Iversen, I.; Kotchoubey, B.; Kubler, A.; Perelmouter, J.; Taub, E.; and Flor. H. (1999). A spelling device for the paralysed. *Nature*, 398:297.298.

Blankertz et al, (2005). BCI Competition III, online: http://www.bbci.de/competition/iii/

Blankertz, B.; Tomioka, R.; Lemm, S.; Kawanabe, M.; and Müller, K-R. (2008). Optimizing spatial filters for robust EEG Analysis, *IEEE Signal Processing Magazine*, pp. 41-56.

Blankertz et al, (2008a). BCI Competition IV, online: http://www.bbci.de/competition/iv/

Blankertz et al., (2008b), BCI Competition IV Results, (submissions by Coyle et al.,), online: http://www.bbci.de/competition/iv/results/index.html

Coyle, D.; Prasad, G.; and McGinnity, T.M. (2004). Improving information transfer rates of a brain-computer interface by self-organising fuzzy neural network-based multi-step-ahead time-series prediction, *Proceedings of the 3rd IEEE Systems, Man and Cybernetics (UK&RI Chapter) conference*, pp. 230-235.

Coyle, D., Prasad, G., and McGinnity, T.M., (2005a) A time-series prediction approach for feature extraction in a brain-computer interface, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 4, pp. 461-467.

Coyle, D.; Prasad, G.; and McGinnity (2005b). A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures, *EURASIP JASP, Trends in Brain-Computer Interfaces (special issue),* vol. 19, pp. 3141-3151.

Coyle, D. (2006) *Intelligent Preprocessing and Feature Extraction Techniques for a Brain Computer Interface*, PhD Thesis, Faculty of Computing and Engineering, University of Ulster, N. Ireland.

Coyle, D.; Prasad, G.; and McGinnity, T.M. (2006a). Creating a nonparametric brain-computer interface with neural time-series prediction preprocessing, *Proc. of the 28th International IEEE Engineering in Medicine and Biology Conference*, pp. 2183-2186.

Coyle, D.; Prasad, G.; and McGinnity (2006b). Enhancing autonomy and computational efficiency of the self-organizing fuzzy neural network for a brain-computer interface, *FUZZ-IEEE, World Congress on Computational Intelligence*, pp. 10485-10492.

Coyle, D.; McGinnity, T.M. and Prasad, G. (2008a) A multi-class brain-computer interface with SOFNN-based prediction preprocessing, *IEEE World Congress on Computational Intelligence*, pp. 3695-3702.

Coyle, D.; Satti, A.; Prasad, G.; and McGinnity, T.M. (2008b). Neural times-series prediction preprocessing meets common spatial patterns in a brain-computer interface, *Proceedings of the 30th International IEEE Engineering in Medicine and Biology Conference*, pp. 2626-2629.

Coyle, D.; Prasad, G.; and McGinnity, T.M. (2009). Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain-computer interface, *IEEE Transactions on Systems*, *Man* and Cybernetics (Part B), vol. 39, issue 6, pp. 1458 - 1471, Dec. 2009.

Davies, D.L. and Bouldin, D.W. (1979). A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 1 No. 4, pp. 224-227.

Dornhege, G.; Blankertz, B.; Krauledat, M.; Losch, F.; Curio, G.; and Müller, K-R. (2006). Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing, *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 11, pp. 2274-2281.

Duda, R.; Hart, P.; and Stork, D. (2001). *Pattern Classification*, 2nd ed. New York: Wiley.

Hassibi, B. and Stork, D. G. (1993). Second order derivatives for network pruning: Optimal brain surgeon, *Advances in Neural Information Processing Systems 4*, pp. 164-171.

Herman, P.; Prasad, G.; McGinnity, T.M.; and Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 16., No. 4, pp. 317-326.

Hornik, K.; Stinchcombe, M.; and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, Vol. 2, pp. 359–366.

Huck, S. W. (2000), *Reading Statistics and Research*. 3rd. ed. New York: Allyn&Bacon/ Longman Pub. Chapter 16.

Iasemidis, L. D. (2003). Epileptic seizure prediction and control, *IEEE Trans. on Biomedical Eng*, vol. 50, no. 5, pp. 549-558.

Jang, J.-S.R., Sun, C. –T., and Mizutani, E. (1997). *Neuro-Fuzzy & Soft Computing*, Englewood Cliffs, NJ: Prentice-Hall, 1997

Kaiser, J.; Perelmouter, J.; Iversen, I.; Neumann, N.; Ghanayim, N.; Hinterberger, T.; Kubler, A.; Kotchoubey, B.; and Birbaumer, N. (2001). Self-initiation of EEG-based communication in paralyzed patients. *Clinical Neurophysiology*, vol. 112, pp. 551–554.

Kasabov, N. K. and Song, Q. (2002). DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction, *IEEE Transactions on Fuzzy Systems,*. vol. 10, no. 2, pp. 144-154.

Kubler, A. ; Kotchoubey, B.; Hinterberger, T.; Ghanayim, N.; Perelmouter, J.; Schauer, M.; Fritsch, C.; Taub, E.; and Birbaumer, N. (1999). The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Exp Brain Res.* vol. 124. pp. 223-232.

Lecuyer, A.; Lotte, F.; Reilly, R. B.; Leeb, R.; Hirose, M.; and Slater, M. (2008). Brain-computer interfaces, virtual reality and videogames, *Computer*, vol. 41, no. 10, pp. 66-71.

Leeb, R.; Lee, F.; Keinrath, C.; Scherer, R.; Bischof, H.; Pfurtscheller, G. (2007). Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 15, pp. 473-482.

Leng, G. (2003). *Algorithmic Developments for Self-Organising Fuzzy Neural Networks*, PhD Thesis, University of Ulster.

Lemm, S.; Schafer, C.; and Curio, G. (2004). BCI competition—Data set III: Probabilistic modelling of sensorimotor μ rhythms for classification of imaginary hand movements, *IEEE Transaction on Biomedical Engineering,* vol. 51, no. 6, pp. 1077-1080.

Mason, S.G.; Bashashati, A.; Fatoruechi, M.; Navarro, K. F.; and Birch, G. E. (2007). A comprehensive survey of brain interface technology designs, *Annals of Biomed. Eng.*, Vol. 35, No. 2, pp. 137-169.

MATLAB® (2009) - http://www.mathworks.com/

McFarland, D. J. and Wolpaw, J. R. (2008). Brain-computer interface operation of robotic and prosthetic devices, *Computer,* vol. 41, no. 10, pp. 52-56.

Owen, A. M. and Coleman, M. R. (2008). Functional neuroimaging of the vegetative state", *Nature Reviews Neuroscience*, Vol. 9, pp. 235-243.

Pfurtscheller, G.; Guger, C.; Muller, G.; Krausz, G.; and Neuper, C. (2000). Brain oscillations control hand orthosis in a tetraplegic, *Neuroscience Letters*, vol. 292, pp. 211–214.

Pfurtscheller, G.; Neuper, C.; Schlogl, A.; and Lugger, K. (1998). Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters, *IEEE Transactions on Rehabilitation Engineering,* vol.6, no.3, pp. 316-324.

Pfurtscheller, G. (1998). *Electroencephalography, Basic Principles, Clinical Application and Related Fields*, 4th Ed., E. Niedermeyer and F. L. Da Silva (Editors), Williams and Wilkins.

Popescu, F.; Fazli, S.; Badower, Y.; Müller, K-R. and Blankertz, B. (2007). Single Trial Classification of Motor Imagination Using Six Dry EEG Electrodes," *PLoS ONE*, vol. 2, 7.

Prasad, G.; McGinnity, T.M.; Leng, G.; and Coyle, D. (2008). On-line identification of self-organizing fuzzy neural networks for modelling time-varying complex systems, *In: Plamen et al. (ed.), Evolving Intelligent Systems*, John Wiley, NY, pp 302-324.

Prasad, G.; Herman, P.; Coyle, D.; McDonough, S.; and Crosbie, J. (2009). Using a motor imagery-based brain-computer interface for post-stroke rehabilitation, *Proc. of the 4th IEEE EMB Conference on Neural Engineering*, pp. 258-262.

Ramouser, H.; Muller-Gerking, J.; and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans. on Rehab. Eng.,* vol. 8, no. 4, pp. 441-446.

Satti, A.; Coyle, D.; and Prasad, G. (2009). Continuous EEG Classification for a Self-paced BCI", *Proc. of the 4th IEEE EMB Conference on Neural Engineering,* pp. 315-318**.**

Satti, A.; Coyle, D.; and Prasad, G. (2008). Optimizing common spatial patterns for a motor imagery-based BCI by eigenvector filtration", *Biomedizinische Technik,* pp. 68-72.

Satti, A.; Coyle, D.; and Prasad, G. (2009). Spatio-spectral & temporal parameter searching using class correlation analysis and particle swarm optimization for a brain-computer interface, *Proceedings of the 2009 IEEE Systems*, Man and Cybernetics Conference, October, 2009.

Silvoni, S.; Volpato, C.; Cavinato, M.; Marchetti, M.; Priftis, K.; Merico, A.; Tonin, P.; Koutsikos, K.; Beverina, F.; and Piccione, F. (2009). P300-based brain–computer interface communication: evaluation and follow-up in amyotrophic lateral sclerosis, *Frontiers in Neuroprosthetics*, Vol. 1, pp. 1-12.

Schlogl et al, (2005a). BCI-Competition III- Dataset IIIa, online: http://www.bbci.de/competition/iii/#data_set_iiia

Schlogl, A.; Lee, F.; Birschof, H.; and Pfurtscheller, G. (2005b) Characterization of four-class motor imagery EEG data for the BCI-competition 2005, *J. of Neural Engineering,* Vol 2, L.14-L.22.

Schlogl, A. ; Keinrath, C.; Zimmermann, D.; Scherer, R.; Leeb, R.; Pfurtscheller, G. (2007b). A fully automated correction method of EOG artifacts in EEG recordings, *Clin. Neurophys.* Vol. 118(1), pp. 98-104.

Schlogl et al, (2008a). BCI-Competition IV- Dataset 2B, online: http://www.bbci.de/competition/iv/#dataset2b

Schlogl et al, (2008b). BCI-Competition IV- Dataset 2A, online: http://www.bbci.de/competition/iv//#dataset2b

Schlogl, A (2009) BIOSIG – an open source software library for biomedical signal processing, online: http://biosig.sourceforge.net/

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control, *IEEE Transactions on Systems, Man and Cybernetics,* vol. 15, no. 1, pp. 116-132.

Tebbens, J.D. and Schlesinger, P. (2006). "Improving Implementation of Linear Discriminant Analysis for the High Dimension/Small Sample Size Problem", *Elsevier Science.*

Wolpaw, J. R.; Birbaumer, N.; McFarland, D. J.; Pfurtscheller, G.; Vaughan, T. M. (2002). Brain-computer interfaces for communication and control, *J. Clinical Neurophysiology*, vol. 113, pp. 767-791.

Wolpaw, J. R. (2004). Brain-computer interfaces for communication and control: Current status, *Proceedings of the 2nd International Brain-Computer Interface Workshop and Training Course, Biomedizinische Technik*, pp. 43-44.

Vaughan, T.M. and Wolpaw, J. R. (2006). Guest Editorial: The Third International Meeting on Brain-Computer Interface Technology: Making a Difference, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2.

Zar, J. H. (1999), *Biostatistical Analysis*. 4th. ed. New-Jersey: Upper Saddle River. p. 255-259.