Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Bioinformatics and Computational Biology (CIBCB 2007)

Channel Selection in EEG-based Prediction of Shoulder/Elbow Movement Intentions involving Stroke Patients: A Computational Approach

Jie Zhou   Sundeep Yedida
Department of Computer Science
Northern Illinois University
DeKalb, IL, USA 60115

***Abstract* – Brain computer Interface (BCI) has gained a lot of attention recently, as a means to detect individuals' intents using brain signals such as electroencephalographic (EEG) for control of machines. In order to achieve the possible use of BCI in stroke rehabilitation, computational intelligent algorithms are important for reliable separation of shoulder versus elbow movement intentions. Efforts have been made on developing data processing and classification algorithm for such task. Differently, this paper investigates the optimal use of electrodes and signal channels, which is formulated as a data-driven feature selection problem. 163 EEG electrodes are used to collect scalp recordings to predict shoulder abduction and elbow flexion intentions in healthy and stroke subjects. We combine the support vector channel selection with a time-frequency synthesized classification algorithm and examine the performances of using different subsets of channel inputs. Preliminary results show that 1) a reduced number of electrodes can be used to achieve the same or better performance than using the full set of signal channels; 2) besides the fact that the accuracy on able-bodied subjects is expectedly higher than the stroke subject, the stroke subject tends to need more electrodes to achieve the best performance; 3) visualization of spatial distribution of channel rankings shows reasonable connection with functional motor cortex areas.**

I. INTRODUCTION

As an approach to detect an individual's intents and to convert brain signals such as electroencephalographic (EEG) into usable control commands, Brain Computer Interface (BCI) has been gaining much attention [1][2]. Recently, there has been study exploring the possibility of applying BCI for clinical use to help patients with moderate to severe motor impairment due to stroke [3][4]. In order to potentially use EEG signals to control a neural prosthesis or other devices for rehabilitation of arm movement discoordination after stroke, it is critical to provide reliable prediction of subjects' motor intentions of shoulder or elbow movements. Effort has been made to explore suitable signal processing and classification algorithms [3][4]. Differently, in this paper, we examine the optimal use of EEG electrodes for the task of classifying mental states for shoulder or elbow movements. In a typical EEG-based BCI study, a number of electrodes are arranged at specific locations of a cap to get scalp EEG recordings. In early studies of BCI, a relatively small number of electrodes are used in the experiments [5][6]. Advancement of multi-channel EEG hardware systems has made it possible to have larger amount of electrodes in recent BCI studies (e.g., 64 electrodes are used in [7] and, in our experiment, scalp recordings were made using a 163-channel EEG system with active electrodes, as shown in Figure 1). Despite that neural-scientists have been using such multi-channel electrode ensemble, most decisions on number of electrodes are based on cost or other ad hoc factors and the determination of the number of channels using computational models has not been systematically explored, especially in the context involving stroke subjects.

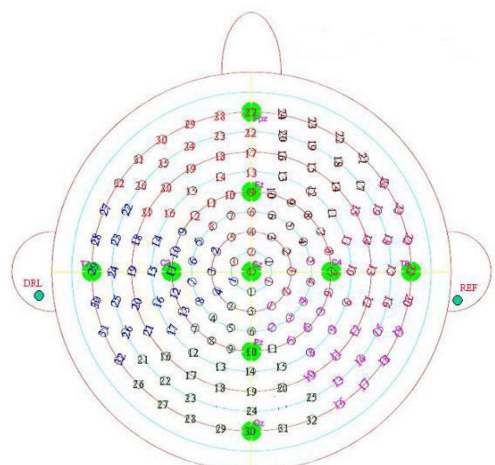


Figure 1 163 electrodes are used for EEG signal recording.

In this paper, we study the channel selection in order to find out input EEG signal channels that are most relevant to the classification task of shoulder abduction and elbow flexion intentions for healthy and stroke patients.

We formulate the problem as a feature selection problem in computational intelligence. The EEG signal channels are thus analogous to features (also called variables in Statistics learning). The input EEG signal is represented by a multidimensional pattern vector $\mathbf{p} = (p_1, \dots p_n)^T$, with n as the number of channels used in the experiment.

Channel selection brings several important advantages to BCI study including:

1) Finding out an optimal number of channels. It is known in the theory and practice of feature selection that multiple features in an input vector typically represent redundancy [8][9]. Feature selection can obtain a subset of features of reduced redundancy yet still retain enough distinguishing information. So in BCI, we can use a subset of channels that are as effective as the full set. In other words, instead of using an unnecessary large number of electrodes, the number of electrodes used in EEG signal collection can be optimized and subsequently the investment on experimental costs can be maximized.

2) Achieving the optimal prediction accuracy. Extending from the above advantage, while redundancy can be reduced by feature selection, we can also select the most relevant features to the prediction task and potentially increase the accuracy compared with using the full set of channels. Such effort can help achieve the best accuracy of classifying mental intentions. It is important for improving the feasibility of applying BCI to clinical use such as stroke rehabilitation since the state-of-the-art EEG-based BCI algorithms for such tasks are yet to be improved.

3) Investigating important EEG signal channels for distinguishing shoulder/elbow movement intentions and their spatial distribution. Physical locations of the important electrodes on scalp can provide feedback on policy of placement of the electrodes. The results on stroke subjects may also contain physiological insights to help neuroscientist and physical therapist to gain more understanding of mechanisms of motor control on healthy and stroke subjects.

In our study, we apply channel ranking and selection to the BCI task by combining the support vector channel selection [10][13] with a weighted time-frequency synthesis classification algorithm [11][4][3]. The approach is data driven with no requirement on prior physiological knowledge. We test the algorithm on two healthy and one stroke subjects. It is the first time that channel selection is investigated for the BCI task of EEG-based classification of shoulder and elbow movement intentions. Same as in [10], we observed that channel selection can reduce the number of channels needed without an increase of error. In addition, we also found that better accuracy may be achieved with an optimal subset. More importantly, preliminary results on the stroke subject indicate that healthy and stroke subjects may require different number of electrodes for optimal performance. The results can potentially have positive impact for developing prostheses for stroke rehabilitation.

II. Data Collection And Signal Preprocessing

We use scalp EEG data collected from two able-bodied (N1, N2) and one stroke subject (S1) at Department of Physical Therapy, Northwestern University, Chicago. The stroke subject has a Fugl-Meyer score of 26/66. Each subject learned to self-initiate the generation of isometric shoulder abduction (SABD) or elbow flexion (EF) at a level of 25% of his/her maximum voluntary torques (MVTs). EEG and torques were collected during the generation of isometric elbow/shoulder torque.

Subjects were cast at the wrist and secured to a six degree of freedom (DOF) load cell with shoulder at 70° abduction angle. The tip of the hand was aligned with the median sagittal plane of the subject and located at a distance from the body which yields an elbow angle of 90°, with 0° representing full extension of the elbow, and the shoulder at approximately a 40° flexion angle. In order to minimize the effect of trunk muscle activation, subjects were seated in a Biodex chair with the trunk secured and the shoulders strapped to the back of the chair. A computer monitor was placed in front of the subject to provide visual feedback of the torque generation during the training protocol.

Scalp recordings were made using a 163-channel EEG system with active electrodes. The electrodes are mounted on a stretchable fabric based on a 10/20 system positioned as illustrated by Figure 1. The cap was fitted on the head of the subject lining the Cz electrode with the intersection of the planes defined by the nasion, inion, and pre-auricular points. EEG data were collected at 1000 Hz sampling rate. Anti-aliasing filtering (100 Hz) was provided before data acquisition. The system was equipped active electrodes that provide a first amplification stage, allowing for the recording of EEG signals with a higher SNR and quicker preparation.

Signals were segmented from -1800ms pre- to -100ms pre-torque onset, and then were baseline corrected and down sampled to 256 Hz. Finite difference surface Laplacian (SL) [14] was applied to signal from each of the channels before exporting the data for further analysis. peripheral electrodes were removed and preprocessed EEG signals from the remaining inner 131 electrodes were exported for BCI classification.

III. Classification Method: Support Vector Enhanced TFSP

For classification of two mental intentions corresponding to shoulder abduction and elbow flexion, we use an improved version of the Time-frequency Synthesized Spatial Patterns algorithm (TFSP).

TFSP is a BCI algorithm developed in [4] and [11] that analyzes event-related desynchronization (ERD) in spontaneous EEG rhythmic activity by decomposing the signal into a time and frequency feature space. In our experiment, a series of 13 frequency bands are obtained, and time intervals of 55ms each are extracted in temporal domain. These time-frequency segments are weighed based on their corresponding contribution and the weights are synthesized during final classification. While the original TFSP algorithm has proved itself an effective feature extraction approach for EEG-based BCI, its prediction power is constrained since a simple correlation comparison with minimal training was used for the classification on the grids. Instead, we use a learning-based classifier to derive weight of time-frequency grid and obtain labels for a new EEG trial on each time-frequency segment. We called our improved method the Classifier-enhanced TFSP. The comparison of the improved version and the original TFSP has found an enhanced accuracy by the new algorithm [3]. In this paper, to facilitate the combination with the channel selection method, we use the support vector classifier [12] as the classifier for the Classifier-enhanced TFSP on each grid for deriving labels and calculating weight of that time-frequency grid.

Support vector classifier is a popular and powerful machine learning algorithm that obtains the weight vector $\mathbf{w}_s$ and offset b of the separating hyperplane between two classes by solving the optimization problem that maximizes the margin, which is defined as the distance between the bounding planes of the two data sets and equivalent to

$$2/\| w_s \|_2.$$

The objective of support vector classifier is thus defined to minimize $\frac{1}{2}\|\mathbf{w}_s\|^2$. The subset of training samples that participate in defining the optimal hyperplane are called "support vectors".

In support vector enhanced TFSP, for each time-frequency grid, its conclusion on the class label d for a testing trial $\boldsymbol{p}(t,f)$ is given by a support vector classifier:

$$d(t,f) = sign(\sum_{i \in SV} \alpha_i l_i < \mathbf{p}_i(t,f), \mathbf{p}(t,f) > +b) \quad (1)$$

where $\boldsymbol{P}_i(t,f)$ is the spatial pattern of ith training trial in the set of support vectors SV; l_i is the desired label of the ith training trial in SV. α_i is the Lagrange coefficient obtained when using Lagrange theory to solve the above optimization problem. A kernel function $k(\mathbf{p}_i(t,f), \mathbf{p}(t,f))$ can replace the inner product in Eq (1) to map samples from the raw data space to a feature space for nonlinear problems. In this paper, linear support vector classifier is used to facilitate the calculation of channel selection criterion (see next Section).

The weight of the time-frequency grid is then obtained by

$$w(t,f) = \begin{cases} [(r(t,f) - Th)/(1-Th)]^4, & r(t,f) > Th \\ 0, & r(t,f) \le Th \end{cases} \quad (2)$$

where $r(t,f)$ is the recognition rate calculated on the grid, and Th is a threshold used to set weights to 0 for grids with low recognition rates. The $w(t,f)$ is not to be confused with the weight vector of support vector hyperplane $\mathbf{w}_s$. A final classification decision for an EEG trial $\mathbf{p}$ is obtained using

$$R(\mathbf{p}) = sign(\sum_{t=1}^{T}\sum_{f=1}^{F} w(t,f) * d(t,f)])$$

In our experiment, R($\mathbf{p}$) = 1 corresponds to the movement intention of shoulder abduction while R($\mathbf{p}$) = -1 indicates an elbow flexion mental intention.

IV. Support Vector Channel Selection

A. Support Vector Channel Selection and Ranking

Many methods can be used to do feature selection [9]. The simplest approach is to assume all features are independent and test them individually based on a selection criterion. However, this approach may not be suitable due to correlations existing among features. Top-down or bottom-up feature selection approaches are more powerful since they can test different subsets of features without assuming features being independent. Meanwhile they avoid the expensive cost of exhaustive testing of all possible permutations.

Support vector channel selection is a top-down approach that uses the Sequential Backward Elimination to select features. The method was first proposed in the context of gene selection for cancer classification [13]. The concept is to determine the importance of a feature/channel based on the influence of the channel has on the margin of a trained support vector classifier. The algorithm starts with the full set of channels. In each round, after training, the weight magnitude $\|\mathbf{w}_s\|$ for each channel is calculated. For linear case, $\mathbf{w}_s$ is calculated as below:

$$w_s = \sum_{i \in SV} \alpha_i l_i \mathbf{p}_i(t,f)$$

The channel(s) with the smallest $\|\mathbf{w}_s\|$ has the lowest influence on the change of objective function and is considered the least important. It is then eliminated from the list of channels. In next round, the signals with the remaining features are trained again using support vector classifier and the channel(s) corresponding to the smallest $\|\mathbf{w}_s\|$ this time is again eliminated. This is done iteratively until there are no channels remaining.

In our experiment, the removed channels are stored in their order of removal in a vector for ranking. To facilitate visualization of important channel groups, ranks of 1 to 10 are assigned to the channels with 10 as the highest rank (most important and last removed from channel list). Each rank is a group of channels. For example, with total of 131 channels, each rank group has 13 channels except the group 1 which has 14 channels.

B. Combining Support Vector Channel Selection with TFSP

We use Support Vector Enhanced TFSP as our BCI algorithm. As explained in Section 3, support vector classifier is applied to every time/frequency grid. So every time/frequency grid calculates the margin and comes up with the candidate channel for elimination. During decision synthesis, a majority voting is conducted among time/frequency grids to determine which channel should be eliminated in current iteration. Only those time/frequency segments with higher than average $w(t,f)$ are eligible voters. If there is a tie on the decision, all tied channels are removed from the set (and added to the rank vector). Table 1 describes the flow of our algorithm of combining support vector channel selection with TFSP.

Table 1. Pseudo code for Support Vector Channel Selection with TFSP.

```
ChannelSet = [1, 2, … , #of channels]
Rank = [ ]
while ChannelSet is not empty
  for t = 1 to # of Time Segments
   for f = 1 to # of Frequency Segments
      apply Support Vector Classifier
      calculate weight vector ws of support vector
hyperplane for each remaining channel in ChannelSet
      get the smallest ||ws|| and the corresponding channel
c(t,f): the candidate channel to be eliminated.
      calculate r(t,f) and w(t,f) for TFSP
    end
  end
  majority voting to determine the channel c to be
eliminated
  remove channel c from ChannelSet.
  add c into Rank
  obtain R(p), the classification decision for testing trials,
and calculate the error rate.
end //of while
```

V. Experiments And Discussions

We applied classified-enhanced TFSP and support vector channel selection to 2 healthy subjects N1 and N2, as well as 1 stroke subject S1. Error rates are reported using 17 fold cross-validation. *Th* in Eq. (2) is set to 0.6. Linear SVM is used for the combined channel selection with classified-enhanced TFSP.

Figure 2 depicts the error rates versus number of channels in the selected set. Polynomial degree-2 fitting is used to get the trend line of the data. Figure 3 displays the distribution of ranking of all channels on the scalp. The warmer color indicates more important channels (e.g. the color red represents rank 10).

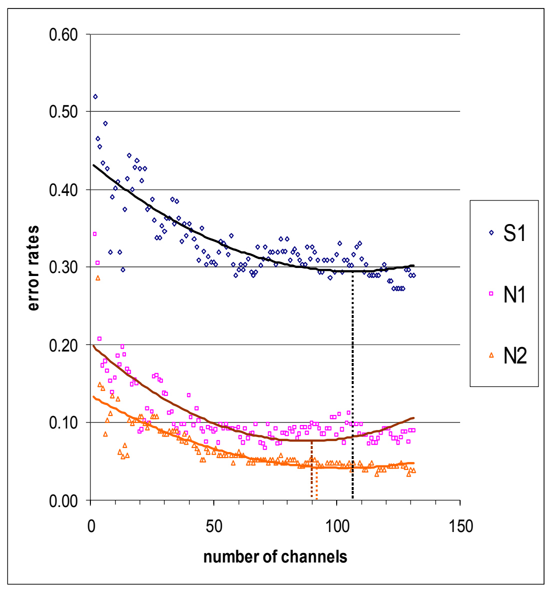
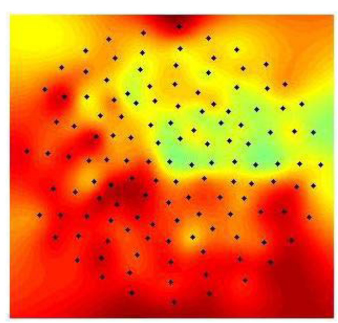
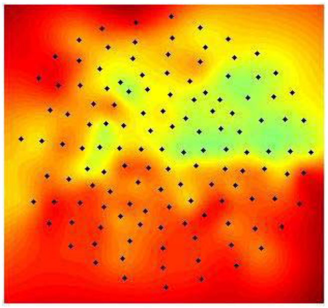


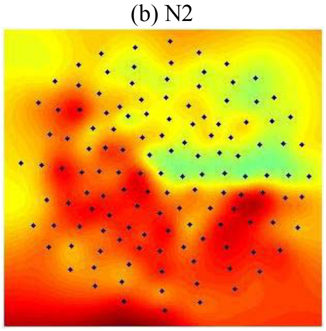
Figure 2. Error Rates vs. Number of Channels.



(a) N1



(b) N2



(c) S1

Figure 3. Rank distribution of channels on scalp. The map is a 2D view when looking down at the top of the head with subject's left-ear to the left. The warm colors indicate higher ranks (important channels). The black dots are the positions of electrodes on the scalp.

From Figure 2, we can see that the proposed classification algorithm is able to achieve accuracies of higher than 90% (10% error rate or lower) on healthy subjects N1 and N2. Expectedly, stroke subject S1 has a higher error rate than two healthy subjects, which may be explained by the possible increase of overlap between cortical activities follow stroke [15]. The trend of error rates versus selected channels shows that, initially, error rates of all three subjects decrease while the number of channels increases, which is a reasonable effect. When the number of channels further increases, we observe the following:

1. For subject N1, the fitting curve shows that error rates actually increase after the number of channels exceeds an optimal point, which is around 90 channels.
2. For subject N2, the fitting curve becomes almost flat when the size of channels reaches certain number, which coincidently is also around 90. This indicates that further increasing number of channels or electrodes beyond this number may not be very effective for reducing error rates.
3. For stroke subject S1, the fitting curve also shows an increase of error rate when the number of channels is big, yet it happens at around 110 channels, which is much later than N1. This suggests that in order to achieve optimal performance, more electrodes are needed for S1 than that of N1 or N2.

In Figure 3, ranks of EEG channels are drawn on the 2D scalp map to show the spatial distribution of the electrodes and their relative significance. The distributions of important channels are consistent among three subjects. Electrodes positioned close to motor cortex area are marked with warm colors, which shows connection to the physiological mechanism despite that the algorithm in this study is purely computational and data-driven. We also hope that such study of channel selection can provide feedback for optimal positioning of electrodes on scalp.

VI. CONCLUSION

This study presents preliminary results of channel selection on using EEG to separate shoulder abduction or elbow flexion intentions of healthy and stroke subjects. Observations confirm the need of exploring important signal channels, not only for the purpose of optimizing cost-effectiveness, but also for enhancing the performance of intention prediction. Current results indicate that healthy and stroke subjects may require different number of electrodes for optimal performance. Future work with more subjects may reveal other differences on optimal electrode usage between healthy and stroke subjects.

ACKNOWLEDGMENT

The authors thank Dr. Jules Dewald at Northwestern University for providing the data for this research. The authors also thank Dr. Jun Yao and Ms. Jie Deng for their helpful discussions.

REFERENCES

[1] J. R. Wolpaw, D. J. MaFarland, T. M. Vaughan, and G. Schalk. "The Wadsworth Center brain-computer interface (BCI) research and development program." *IEEE Trans. Neural. Syst. Eng.* 11:204-207, 2003.
[2] A. Vallabhaneni, T. Wang T and B. He. Brain-computer interface. *Neural Engineering* ed B He (New York: Kluwer/Plenum) 85-122, 2005.
[3] J. Zhou, J. Yao, J. Deng, and J. Dewald. "EEG-based Discrimination of Elbow/Shoulder Torques using Brain Computer Interface Algorithms: Implications for Rehabilitation". *Proc. 27th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* 2005. pp4134 - 4137
[4] J. Deng, J. Yao and J. Dewald, "Classification of the intent to generate a shoulder versus elbow torque by means of a time-frequency synthesized spatial patterns BCI algorithm," *J Neural Engineering*, 2:131-138, 2005.
[5] G. Pfurtscheller, C. Neuper, A. Schlogl, K. Lugger, "Separability of EEG signals recorded during right and left

motor imagery using adaptive autoregressive parameters," IEEE Trans. Rehabil. Eng, 6(3), pp316-325, 1998

[6] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C.A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol,* 78, pp252-259, 1991.

[7] A. Osman and R. Albert, "Time-course of cortical activation during overt and imagined movements," Cognitive Neuroscience Annual Meeting, New York.

[8] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans on Pattern Analysis and Machine Intelligence,* 27(8), pp.1226-1238, 2005.

[9] A. Webb. *Statistical Pattern Recognition*. John Wiley and Sons. 2002.

[10] T. N. Lal, M. Schrüder, T. Hingerberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf. "Support Vector Channel Selection in BCI". IEEE Transactions on Biomedical Eng. 51(6):1003 – 1010, 2004.

[11] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns". *Clinical Neurophysiology,* 115, 2744–2753, 2004.

[12] V. Vapnik. *The nature of statistical learning theory.* Springer-Verlag. New York. 1996.

[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *J. Machine Learning Res.*, vol. 3, pp. 1439–1461, March 2003.

[14] B. Hjorth, "An on-line transformation of EEG scalp potentials into orthogonal source derivations," *Electroencephalogr. Clin. Neurophysiol.* 39, pp526-530, 1975.

[15] J. Yao, M. Ellis, and J. Dewald, "Mechanism and rehabilitation of discoordination following stroke using a cortical imaging method," *Proc. of 27th Annual Int. Conf. Of the IEEE Engineering in Medicine and Biology Society, 2005.*