Proceedings of the 4th International
IEEE EMBS Conference on Neural Engineering
Antalya, Turkey, April 29 - May 2, 2009

EEG Signal Classification during Listening to Native and Foreign Languages Songs

Shao-Jie Shi[1] and Bao-Liang Lu[1,2]
[1]Department of Computer Science and Engineering
[2]MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems
Shanghai Jiao Tong University
800 Dong Chuan Rd., Shanghai 200240, China
bllu@sjtu.edu.cn

Abstract—**This paper designs an experiment to analyze different EEG patterns while subjects are listening to different language songs. In the process of experiment, the subjects listen to multi-section songs. Every two songs have the same rhythm and only the lyrics are different, one in Chinese and the other in Japanese. The songs are sung by one singer and the Chinese subject don't know Japanese at all. At the same time we collect the EEG signals which are supposed to have very subtle difference corresponding to two kinds of songs. Then we use common spatial pattern algorithm to extract features and define an average energy function to represent them. After that we use support vector machine to learn and classify the EEG data. We find that the difference pattern mainly lay in low spectral band (0-0.5Hz), and concentrate on the left frontal area of the cortical. We achieve the highest classification accuracy of 97.30% and an average classification accuracy of 87.15%.**

I. Introduction

We get most of our knowledge through seeing and listening. This paper devotes to the research of our response to the music stimuli. We receive information by listening everyday, such as speech, music, nature voice and so on. Our brain will response differently to different sound. For example, when we listen to native language and foreign languages, our brain will response differently [1]. How our brain response to these and how to measure it is still under study. In this paper, we try to use EEG signals to analyze the different mental states when we are listening to songs with different language lyrics.

Regions involved in language processing have been observed in the inferior part of the left temporal lobe. According to Pallier's study [1], for each subject, the brain region where activation due to native language was significantly stronger than that due to foreign language. There was a large activation of left-lateralized temporal and inferior frontal regions when subjects listening to native sentences. It is anticipated that a substantial increase in gamma-band activity would be observed during visual word perception [2]. David and colleagues indicated that verbal learning with a musical template strengthens coherent oscillations in frontal cortical networks involved in verbal encoding [3]. Lin and colleagues used MLPs to classify four classes of EEG data when the subject was listening to four kinds of music (joy, angry, sadness, pleasure), and achieved the average classification accuracy of 69.69% [4]. We can see there exists differences in EEG when we are listening to different languages songs.

TABLE I
THE DISPLAY ORDER OF SONG SECTION

0	1st Section		2nd Section		...	22nd Section	
10s	Chinese song	10s	Japanese song	10s	...	Japanese song	10s

The purpose of this study is to quantify the different patterns of EEG signals using machine learning methods when the subjects are listening to different language songs and we also try to explain the difference and find more prior knowledge from the physiological aspect. This paper designs a new experiment which discovers the different patterns when people are listening to different languages songs. We find that these differences mainly lay in 0-3.5Hz and γ (26-50Hz) rhythm and the place of these difference concentrate on the left frontal region which is consistent with the result of fMRI [1]. We use Common Spatial Pattern algorithm [5] [6] and support vector machines (SVMs) to learn and classify the EEG data.

II. METHODS

A. Experiment Setup

In the experiment, we choose two kinds of songs (Chinese lyrics and Japanese lyrics) to stimulate the subject and at the same time we record the EEG signals using 64-channel Neuro-scan device. All songs are sung by one singer and the same song is sung in two versions, namely Chinese lyrics and Japanese lyrics. The number of song sections is 22 which has 11 Chinese songs and 11 Japanese songs. The length of song section ranges from 40 seconds to 80 seconds. The Chinese subjects can't understand Japanese at all and didn't receive professional music training before.

Table 1 is about the display order of songs. There are 10 seconds between two song sections for rest. During the process of experiment, the subject sits in a comfortable chair and the sound box will play the song. Then we record the EEG signals through extended 10/20 system with Neuro-scan cap. The experiment lasts for about 25 minutes.

B. Data preprocessing

Firstly, we resample the data with rate of 100Hz; then we divide the original signal to 22 sections according to the number of songs; after that we divide each section to many data samples every 4 seconds, namely 400 points for a sample according to 100Hz sampling rate. Here we would like to mention that we also try some other number of points per sample and the classification results are showed in Fig. 2. But when the number of points increases, the number of training samples will decrease. As a result, the classification precision will decrease. The experimental results show that 400 points per second is a compromise between highest classification precision and sample number; finally we divide the samples to training data and testing data using the ratio of 3:1. We try to make the training data scattered to cover the whole data which is proved to be effective. We label Chinese songs as Class A, and Japanese songs as Class B. We totally get 6 persons' data and each person's experiment time is about 25 minutes.

C. Common Spatial Pattern

Common spacial pattern (CSP) algorithm is the simplification of Common Spatial Subspace Decomposition (CSSD) [7] algorithm and is designed according to the theory of simultaneous diagonalization of the covariance matrices of two classes [9]. The main idea is to use a linear transform to project the multi-channel EEG data into low-dimensional spatial subspace with a projection matrix. So the essence is to compute a projection matrix of two classes' data. After the projection, one signal will have the maximum variant and the other signal will have the smallest variant at the same direction and the sum of the two variant will be 1. CSP is a supervised learning method and is very effective on two-class problem, because it use the class information to extract the feature which has the biggest difference between two classes.

Suppose that X_a and X_b are $N \times T$ signal matrices. In our experiment, X_a represents the EEG signal when subject is listening to Chinese songs and is concatenated by all the processed training sample of class A. X_b represents the EEG signal when subject is listening to Japanese songs. Suppose that N is the channel number and T is the number of all training sample points of one class.

The estimation of covariant matrix can be expressed as

$$R_a = (X_a \cdot X_a^T)/(T-1) \tag{1}$$

$$R_b = (X_b \cdot X_b^T)/(T-1) \tag{2}$$

$$R = R_a + R_b = U_0 \cdot \Sigma \cdot U_0^T \tag{3}$$

Whitening transformation matrix P can be expressed as

$$P = \Sigma^{-1/2} \cdot U_0^T \tag{4}$$

By whitening them, we have

$$S_a = P \cdot R_a \cdot P^T \tag{5}$$

$$S_b = P \cdot R_b \cdot P^T \tag{6}$$

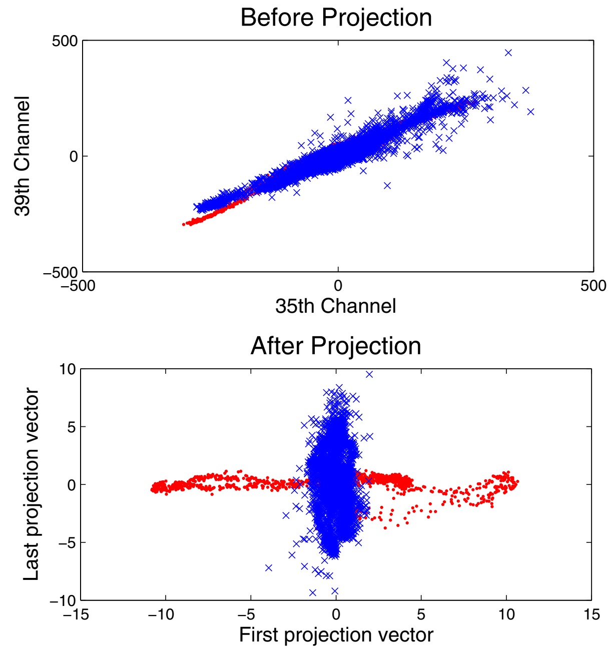

Fig. 1. CSP projection of the original EEG signal. Before projection, we show the data point of two channels which are completely mixed. After projection, the points from class A (red points) have the maximum variant in the x axis and the smallest variant in the y axis.

where

$$S_a + S_b = I \tag{7}$$

By diagonalization, we have

$$S_a = U \cdot \Sigma_a \cdot U^T \tag{8}$$

$$S_b = U \cdot \Sigma_b \cdot U^T \tag{9}$$

where

$$\Sigma_a + \Sigma_b = I \tag{10}$$

Finally we get the projection matrix as follows:

$$SF = U^T \cdot P \tag{11}$$

where SF is the projection matrix and is called Spatial Filter. Each row of SF matrix corresponds to a projection vector. In the direction of first projection vector, the projection of the data from Class A will have maximum variant and the Class B will have the smallest variant. The sum of the two variants is 1. This is the basic idea of the CSP algorithm.

After projection, we have

$$cov(SF \cdot X_a) + cov(SF \cdot X_b) = I \tag{12}$$

where

$$cov(X) = (X \cdot X^T)/(T-1) \tag{13}$$

We can see from Fig. 1 that before projection two signals were completely mixed, and after projection they distinguish from each other according to the CSP algorithm.

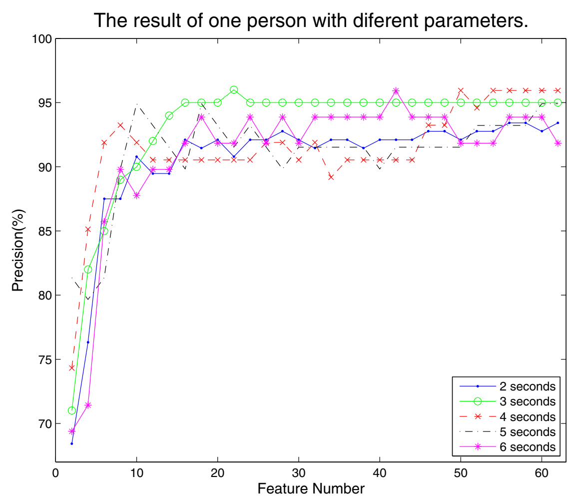

Fig. 2. The result of subject 2 with different sample lengths from 2 seconds to 6 seconds.

D. Feature organization

Here we have the projection matrix SF and many processed sample date $T_{62\times400}$; then we project each sample data $T_{62\times400}$ using the first and last M rows of SF matrix SF^{2M}, we get a matrix $K_{2M\times400}$; finally we define an average energy function to calculate the average energy of 400 points data and get a $2M \times 1$ vector v which is the feature of this sample.

This feature organization process can be described as follows.

By projection, we have

$$K = SF^{2M} \cdot T \tag{14}$$

The feature for one sample is defined as

$$v = (v_1, v_2 ... v_{2M})^T \tag{15}$$

The average energy function is defined as

$$v_i = \frac{1}{400} \sum_{j=1}^{400} K_{ij}^2 \tag{16}$$

After projection, we assume each sample data has the same distribution of the whole data set of its class. Then in the first projection direction, namely the first row of SF matrix, the sample belong to Class A will have the biggest variant. Here we use Average Energy Function which includes both the variant and mean information of the sample data. This can be seen from the following equation:

$$E(v_i^2) = var(v_i) + E^2(v_i) \tag{17}$$

E. SVM classification

We use LIBSVM [8] as classifiers and choose the RBF as kernel to learn and classify the EEG data. SVM has two main parameters, gamma and c, which need to be grid searched for the best. Here we choose [0.001 0.01 0.05 0.1 1 10] as the candidate values of gamma and [1 10 50 100 300 500 1000] as the candidate values of c. Then for each person's data, we use the Brute-force method to calculate the best parameters which produce the highest classification precision.

TABLE II
THIS TABLE SHOWS THE BEST RESULTS OF 6 SUBJECTS IN ALL SPECTRAL BANDS. THE LENGTH OF EACH SAMPLE IS 4 SECONDS. WE TOTALLY GET 228 TRAINING SAMPLES AND 76 TESTING SAMPLES PER SUBJECT. IN TRAINING SAMPLES, WE HAVE 116 SAMPLES OF CLASS A AND 112 SAMPLES OF CLASS B.

Subject	Precision(%)	Recall(%)	F1(%)
S1	86.05	86.05	86.05
S2	96.54	96.51	96.52
S3	86.12	86.05	86.09
S4	87.34	87.38	87.36
S5	93.13	93.17	93.15
S6	82.77	82.70	82.73
Average	88.66	88.65	88.65

TABLE III
THE TEST ACCURACY OF 6 SUBJECTS IN DIFFERENT RANGE OF SPECTRAL. "ALL" REPRESENTS THAT ALL THE FEATURES IN DIFFERENT SPECTRAL BANDS ARE USED. THE LENGTH OF EACH SAMPLE IS 4 SECONDS.

Subject	0-0.5Hz (%)	δ (%)	θ (%)	α (%)	β (%)	γ (%)	All (%)
S1	82.43	72.97	66.22	71.62	71.62	71.62	82.43
S2	97.30	85.14	68.92	72.97	79.73	78.38	97.30
S3	83.78	75.68	74.32	81.08	87.84	91.89	89.19
S4	86.84	68.42	64.47	63.16	61.84	64.47	85.53
S5	90.79	73.68	63.16	75.00	69.74	68.42	90.79
S6	75.00	68.42	67.11	65.79	67.11	72.37	77.63
Average	86.02	74.05	67.37	71.60	72.98	74.52	87.15

III. RESULT AND DISCUSSION

The experiment results are showed in Tables II and III, and Figs. 2 and 3.

In Table II, we choose the parameters which have the best performance. From this table, we can see that EEG signal differs from one person to another and the subject 2 achieved the highest performance over most parameters. And these results are achieved in all spectral bands. The Precision, Recall and F1 here are the evaluation methods usually defined in text classification.

Figure 2 illustrates the results of subject 2 with different parameters including the number of features and sample length. We can see from this figure that the number of features is the most important parameter in this experiment. In this experiment, we used the whole training set to compute the projection matrix SF and then we project each sample with SF. In the idea situation, for class A, v_1 will be the biggest and v_{2M} be smallest feature. Only two features will distinguish the data. But the actual situation is that mental state will be different even when one is listening to the same song.

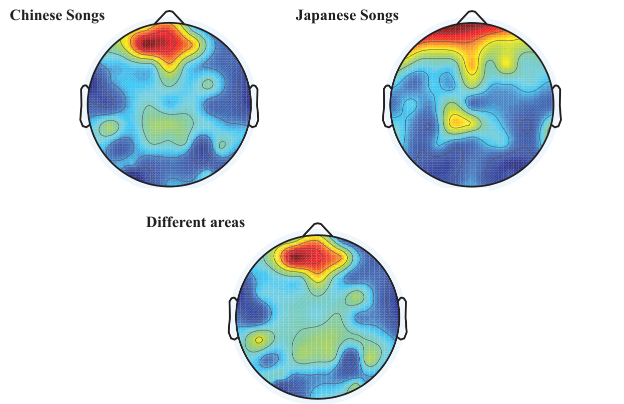

Fig. 3. Scalp topographies of EEG energy proportion in 0-0.5Hz. The top-left scalp represents the state when people are listening to Chinese songs and the top-right scalp Japanese songs. The bottom scalp is the difference of two scalps which is calculated by the absolute value of the difference between the top-left scalp energy and the top-right scalp energy.

So increasing the number of features would increase the information and are supposed to reduce the classification error. We can see from Fig. 2 that when the number of features is more than 20, we can achieve higher classification accuracy.

Table III is about the results of subjects in different spectral bands. We use low-pass, band-pass, and high-pass FIR filters to process the data. The length of sample is 4 seconds. Here Walter's definition of spectral bands, namely δ (0.5-3.5Hz), θ (4-7Hz), α (8-13Hz), β (14-25Hz) and γ (26-50Hz), are introduced. We find that we achieve the highest performance in the lower spectral band (0-0.5Hz) due to the effect of the Slow Cortical Potentials (SCP) [10] [11]. Here accuracy is defined as the percentage of the right classified test samples. For subject 2, we get the highest classification accuracy of 97.30%. We also find that classification accuracy improve a little in β and γ rhythms, especially for subject 3 who achieved the highest performance in γ rhythm. Finally, we combine all the features in different spectral bands and obtain an average improvement of 1.13% in accuracy.

Since we achieved higher performance in low spectral band, we use the FIR band-pass filter (0-0.5Hz) to process two classes EEG signals, separately, and then we use the periodogram [12] to estimate the Power Spectral Density of each channel. We plot the energy between 0 and 0.5Hz of each channel in the scalp. The results are shown in Fig. 3. We find that the typical different area is near the left frontal region of the scalp, which is consistent with the result of fMIR [1].

IV. Conclusions

In this paper, we have designed a new experiment based on different languages songs stimuli, namely Chinese songs which can be understood by the native subjects and Japanese songs which the subjects don't understand at all. We recorded the EEG signals using extended 10-20 Neuro-Scan system. Finally, we totally collected six subjects' EEG data for the analysis.

From the analysis of the EEG data, we can see that when people are listening to different languages songs, the brain will produce different EEG signal patterns, which lay in 0-0.5Hz and mainly in the left frontal region of the brain. We used common spatial pattern to project the original data and define an average energy function to represent the features. Then we used SVM classifiers to classify the data. Finally, we achieved the highest classification accuracy of 97.30% and an average classification precision of 87.15% for two different languages songs. The results of the experiment confirm in a different way that native language and foreign language will cause different reaction in left frontal brain region [1].

Acknowledgments

This work was supported in part by the National Basic Research Program of China under the grant 2009CB320900 and Natural Science Foundation of China under the grants NSFC 60773090 and NSFC 90820018. We thank Li-Chen Shi, Jia-Wei Fu, Mu Li and Xiao-Wei Wang for the discussion of the experiments and this paper.

References

[1] Pallier, C. et al, "Brain imaging of language plasticity in adopted adults: can a second language replace the first?", *Cereb. Cortex*, vol. 13, no. 2, pp.155-161, 2003.

[2] Kyung Hwan Kim, Ja Hyun Kim, "Analysis of induced gamma-band activity in EEG during visual perception of Korean, English, Chinese words", *Neuroscience Letters*, vol. 403, no. 3, pp.216-221, Aug. 7, 2006.

[3] Peterson, David A., and Michael H. Thaut, "Music increases frontal EEG coherence during verbal learning", *Neuroscience Letters*, vol. 412, no. 3, pp. 217-221, February 2007.

[4] Y. Lin, C. Wang, T. Wu, S. Jeng and J. Chen, "Multilayer Perceptron for EEG Signal Classification during Listening to Emotional Music", *TENCON 2007 - 2007 IEEE Region 10 Conference*, vols. 1-3, pp. 236-238, 2007.

[5] H. Ramoser, J. M. Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Trans.Rehab.Eng.*, vol. 8, no. 4, pp. 441-446, 2000.

[6] Yijun Wang, Shangkai Gao, Xiaorong Gao, "Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface", *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, Sep.1-4, 2005.

[7] Y. Wang, P. Bergc, M. Scherga, "Common Spatial Subspace Decomposition Applied to Analysis of Brain Responses under Multiple Task Conditions: A Simulation Study", *Clinical Neurophysiology*, vol. 110, pp. 604-614, 1999.

[8] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines", Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.

[9] Fukunaga, Keinosuke, "Introduction to Statistical Pattern Recognition. 2nd ed", Academic Press Prof., Inc., San Diego, CA, 1990.

[10] Mensh B.D., Werfel J., Seung H.S, "BCI competition 2003-data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of Electroencephalographic signals", *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1052-1056, 2004.

[11] N. Birbaumer, T. Elbert, A. G. Canavan,and B. Rockstroh, "Slow potentials of the cerebral cortex and behavior", *Physiol. Rev.*, vol. 70, pp. 1-41, 1990.

[12] A. V. Oppenheim, R. W. Schafer, "Digital Signal Processing", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.