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
blu@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.

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 $\gamma$ (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.

<table>
<thead>
<tr>
<th>0</th>
<th>1st Section</th>
<th>2nd Section</th>
<th>...</th>
<th>22nd Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>10s</td>
<td>Chinese song</td>
<td>10s</td>
<td>Japanese song</td>
<td>10s</td>
</tr>
</tbody>
</table>
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 spatial 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 of first projection vector, the projection of the data point of two channels which are completely mixed. After projection, we have

\[
\begin{align*}
S_a + S_b &= I \\
S_a &= U \cdot \Sigma_a \cdot U^T \\
S_b &= U \cdot \Sigma_b \cdot U^T
\end{align*}
\]

where

\[
S_a + S_b = I
\]

By diagonalization, we have

\[
S_a = U \cdot \Sigma_a \cdot U^T \\
S_b = U \cdot \Sigma_b \cdot U^T
\]

where

\[
\Sigma_a + \Sigma_b = I
\]

Finally we get the projection matrix as follows:

\[
SF = U^T \cdot P
\]

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

\[
cov(SF \cdot X_a) + cov(SF \cdot X_b) = I
\]

where

\[
cov(X) = (X \cdot X^T)/(T - 1)
\]

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.
D. Feature organization

Here we have the projection matrix $SF$ and many processed sample data $T_{62 \times 400}$; then we project each sample data $T_{62 \times 400}$ using the first and last M rows of $SF$ matrix $SF^{2M}$, we get a matrix $K_{2M \times 400}$; 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

$$K = SF^{2M} \cdot T \quad (14)$$

The feature for one sample is defined as

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

The average energy function is defined as

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

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:

$$E(v_i^2) = var(v_i) + E^2(v_i) \quad (17)$$

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.

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.
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