INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Classify Imaginary Mandarin Tones with Cortical EEG Signals

Hua Li [1, 2] *, Fei Chen* [2]

[1] Shenzhen University Health Science Center School of Biomedical Engineering, Shenzhen, China
[2] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China
fchen@sustech.edu.cn

Abstract

Speech synthesis system based on non-invasive brain-computer interface technology has the potential to restore communication abilities to patients with communication disorders. To this end, electroencephalogram (EEG) based speech imagery technology is fast evolving largely due to its advantages of simple implementation and low dependence on external stimuli. This work studied possible factors accounting for the classification accuracies of EEG-based imaginary Mandarin tones, which has significance to the development of BCI-based Mandarin speech synthesis system. Specially, a Mandarin tone imagery experiment was designed, and this work studied the effects of electrode configuration and tone cuing on accurately classifying four Mandarin tones from cortical EEG signals. Results showed that the involvement of more activated brain regions (i.e., Broca's area, Wernicke's area, and primary motor cortex) provided a more accurate classification of imaginary Mandarin tones than that of one specific region. At the tone cue stage, using audio-visual stimuli led to a much stronger and more separable activation of brain regions than using visual-only stimuli. In addition, the classification accuracies of tone 1 and tone 4 were significantly higher than those of tone 2 and tone 3.

Index Terms: Electroencephalogram (EEG), speech imagery, Mandarin tones, support vector machine (SVM).

1. Introduction

Speech synthesis technology based on pronunciation imagination is a process of making a machine to produce understandable speech by utilizing the psychological, physiological and physical characteristics of human being when imagining speech pronunciation [1-2]. Traditional speech synthesis technology models the physical structure of human voice to produce speech (e.g., tongue motion, airflow variation) [e.g., 3]. The parameters of this kind of system are too many and complex to be optimized. Subsequently, a trainable speech synthesis system combining big data was developed [e.g., 4]. Through the training of a larger number of data, this kind of system obtained the statistical acoustic model of speech, and then generated the corresponding speech synthesizer. Although this system constructed an acoustic model automatically and improved the effect of speech synthesis, it had some problems (e.g., poor model generalization ability), and commonly needed a lot of annotated data. The above two methods are mainly based on the physical information of speech pronunciation for modeling analysis, ignoring the psychological and physiological characteristics for speech production. Studies are actively ongoing towards reconstructing understandable speech in tasks of speech imagery. Recently, based on the neural electrical signal information, Chang et al. fused the pronunciation model parameters and reconstructed the comprehensible speech information [1]. Hong et al. used electroencephalogram (EEG) information to reconstruct Chinese phonetic information [2]. Herff and Schultz used electrocorticogram (ECoG) data as the input signal of automatic speech recognition technology and successfully realized the recognition of audibly spoken speech [5]. While these studies showed that brain-computer interface (BCI) technology had become a new development direction of intelligent speech synthesis system, the EEG signals in these studies were largely recorded by invasive electrodes, making them difficult to be applied in real situations. Meanwhile, for reconstructing an understandable speech from speech imagery tasks, a lot of work has to been done to explore the neuropsychological principles of speech imagery tasks [e.g., 6-10]. For instance, Tian et al. explored the neuropsychological principles of people performing speech imagery tasks by using two methods of speech imagery (i.e., speaking imagery and hearing imagery), and compared the activation levels and regions of the brain under the two methods [6-7]. Matsumoto et al. studied speech imagery to classify two of five Japanese vowels, and achieved a classification accuracy of 77% [8].

Most existing speech imagery tasks focus on the perception and classification of non-tonal languages, e.g., primarily the recognition of vowels, consonants or words in English. As a tonal language, Mandarin has four lexical tones, i.e., the flat tone, the rising tone, the falling-rising tone, and the falling tone (usually expressed as tone 1, tone 2, tone 3 and tone 4) [11-12]. Lexical tone plays an important role in the recognition and understanding of Chinese characters [13], and the neuropsychological principles of Chinese speech imagery tasks are different from those of non-tonal languages [14-15]. To date, little work has been done to investigate the neuropsychological principles of Mandarin tone imagery.

The aim of this work was to study the classification of imaginary Mandarin tones with cortical EEG signals. A Mandarin tone imagery experiment was designed, including two types of tone cuing methods (i.e., with visual stimuli and audio-visual stimuli) before the tone imagination stage. The possible accounts for the tone classification accuracy were analyzed, including the effects of EEG electrode configuration used in tone classification and tone cuing to activate brain regions before Mandarin tone imagination.

http://dx.doi.org/10.21437/Interspeech.2020-1248

2. Methods

2.1. Subjects and materials

Fourteen (6 male and 8 female) Mandarin-speaking students from Southern University of Science and Technology took part in the experiment. All participants were physically healthy and had no neurological or psychological problems. They signed informed consent forms and were paid for their participations. The procedure involving human experiment was reviewed and approved by the Research Ethics Committee of the Southern University of Science and Technology.

Two types of stimuli were used to cue Mandarin tone imagination, yielding two conditions. The first condition used visual-only stimuli, while the second condition used combined audio-visual stimuli. Figure 1 shows the visual stimuli presented as pictures used in the first condition (i.e., visual-only). In order to control the interference of image transformation, all symbols/signs in the pictures were presented at the same size. The visual stimuli in condition 2 (i.e., combined audio-visual) were the same as those used in condition 1, while the audio stimuli in condition 2 were syllable /ba/ in four Mandarin tones pronounced by an adult female native Mandarin-Chinese speaker. The pronounced Mandarin syllables were recorded at a sampling rate of 16 kHz, and their durations were normalized to 1 sec.

2.2. Experimental paradigm and procedure

Each stimulus condition consisted of 4 test blocks (the order of stimuli presentation was pseudo-random), and subjects were asked to imagine 40 Mandarin tones in each block. Each speech imagery trial was divided into a tone cue stage (with visual-only stimuli or audio-visual stimuli) and a subsequent tone imagination stage. Subjects pressed any key to start the experiment when they were ready. The visual-only stimuli (or combined audio-visual stimuli) appeared for 1 sec at the tone cue stage. Then at the tone imagination stage (immediately following the tone cue stage), a prompt sign '+' appeared to inform the participant to imagine the corresponding tone. The tone imagination stage lasted for 2 sec. The total duration of one stimulus condition was around 8 mins. All audio stimuli were presented through a Sennheiser HD 25 earphone at a comfortable level. During the whole experiment, participants sat in an acoustically and electrically shielded chamber. They were seated comfortably and instructed to pay attention to visual-only (or combined audio-visual) stimuli.

2.3. EEG data recording and pre-processing

The EEG signals were recorded through a 64-channel electrode cap (Neuroscience Inc.). Following the extended international 10-20 system, the cap was placed at specific positions. The top of the nose served as a reference for all electrodes, and the ground electrode was attached to the forehead. The impedance between the reference electrode and any recording electrode was less than 5 kΩ. The sampling rate of EEG data was 500 Hz. To reduce unnecessary motion artifacts, the subjects were asked to minimize the body movements. The EEG data at the tone cue stage and the tone imagination stage were recorded.

The recorded EEG signals were analyzed with EEGLAB 14.1.1. First, the data were re-referenced using the contralateral mastoid signals. Epochs containing artifacts

Figure 1: *Visual stimuli used in this experiment, including Chinese syllable /ba/ in 4 lexical tones and the sign '+' that prompts the subjects to imagine the Mandarin tone.*

exceeding ±75 μV were excluded from the averaging procedure. Then the artifact (e.g., eye blinks, horizontal eye movement, electrocardiographic activity) correction was performed by infomax independent component analysis. After artifact removal, for the EEG signals at the tone cue stage, each epoch was selected between 100 ms pre-stimulus and 1000 ms post-stimulus and corrected with the baseline of the pre-stimulus time window; for the EEG signals at the tone imagination stage, the range of each epoch was from -100 to 900 ms, and the baseline was corrected with that of the pre-stimulus time window.

2.4. EEG data processing

For the EEG data recorded at the tone cue stage, the EEGLAB toolbox was used to calculate the peak amplitude and latency. For the EEG signals at the tone imagination stage, the common spatial patterns (CSP) were used to extract the features of tone pronunciation imagination from the EEG data, and then an adaptive support vector machine (SVM) was used to classify and recognize Mandarin tones.

For EEG data processing, CSP is widely noted as one of feature extraction methods with the best performance and the most extensive applications [e.g., 16-19]. The CSP processing extracts EEG features by maximizing the variance of the data between classes and minimizing the variance within classes. Since CSP is only applicable to dichotomies, the following method was used to extract features for the problem of four tone classification in this work. If tone 1 was treated as one type of data, the other three tones were treated as another type of data. Based on these two types of data, the characteristics of tone 1 could be extracted, and the characteristics of the remaining three tones were further extracted [20].

SVM has strong generalization ability and strong adaptability to the problem of over-fitting and curse-of-dimensionality [21]. It has been widely used to classify the EEG data of speech imagery, and better classification accuracies were obtained in early studies [e.g., 16, 22-23]. The basic principle of SVM is to find an optimal classification hyperplane to maximize the classification interval between two types of data. In this work, several sub-classifiers were trained to solve the 4-class classification task. First, one classifier was trained for any pair of 2 classes chosen from the four tone classes, and the total number of 2-class classifiers was six (i.e., tone 1 vs. tone 2, tone 1 vs. tone 3, tone 1 vs. tone 4, tone 2 vs. tone 3, tone 2 vs. tone 4, and tone 3 vs. tone 4). Then, the EEG data to be classified were processed by all 2-class classifiers, and a vote was used to determine the final class attribute of the Mandarin tone.

It is feasible to use all 64 channels of EEG data to categorize tones. This work studied the effect of different language-relevant brain regions on Mandarin tone imagination and classification. For the speech imagery task, the brain's main regions of activation were Broca's area, Wernicke's area, and the primary motor cortex [12]. Each area had EEG signals

Table 1. *Descriptive statistics (average ± standard deviant) of the classification accuracies of 4 Mandarin tones at different electrode configurations. Broca's area: electrodes F5 and FC3, Wernicke's area: electrodes P5 and CP3, and primary motor cortex: electrodes C3 and C4.*

Conditions	**2 electrodes**			**All 6 electrodes** (%)
	Broca's area (%)	Wernicke's area (%)	primary motor cortex (%)	
Visual-only	39.9 ± 1.3	31.6 ± 1.1	34.4 ± 1.2	67.7 ± 1.0
Combined audio-visual	55.2 ± 1.1	39.2 ± 1.0	41.2 ± 0.9	80.1 ± 1.2

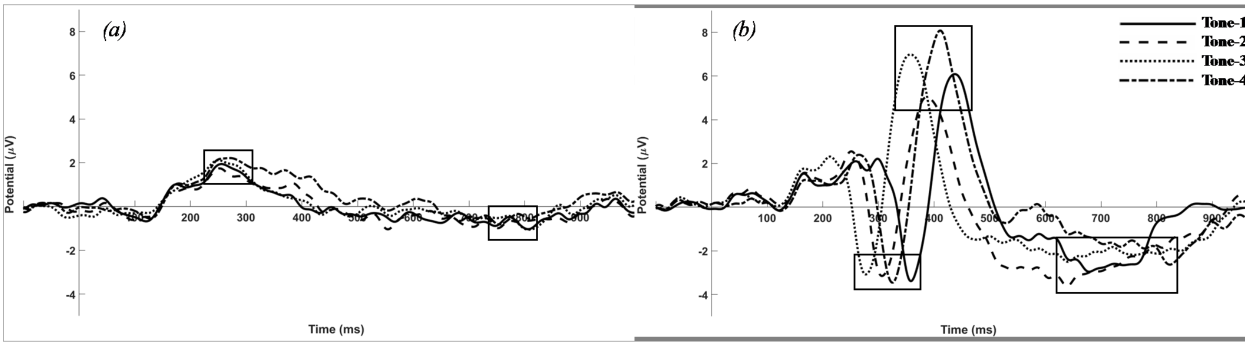

Figure 2: *The average brainwaves of all subjects at electrode CZ at the tone cue stage. (a): Average brainwaves under the visual-only condition; and (b): Average brainwaves under the combined audio-visual condition. The rectangle represents the area where the peak occurs for different tones.*

recorded from 2 electrodes (see Table 1) in this work, i.e., F5 and FC3 at Broca's area, P5 and CP3 at Wernicke's area, and C3 and C4 at primary motor cortex. Then, this work also used the EEG signals recorded from the 6 electrodes involved in those three brain regions to classify Mandarin tones, aiming to achieve a good accuracy of tone classification under the premise of reducing the amount of data.

3. Results

3.1. Effect of electrode configuration on classification accuracy

Table 1 shows the average classification accuracies (across all 4 Mandarin tones) for 4 types of electrode configurations under the 2 tone cuing conditions (i.e., visual-only and combined audio-visual). It is seen that when the EEG signals from 2 electrodes at a specific brain region are utilized to classify imaginary Mandarin tones, the EEG information from Broca's area yields the best classification accuracy under both tone cuing conditions, i.e., 55.2% under the combined audio-visual condition and 39.9% under the visual-only condition. In addition, when EEG signals from all 6 electrodes at Broca's area, Wernicke's area and primary motor cortex are used to classify imaginary Mandarin tones, the classification accuracies are significantly improved (p<0.01), i.e., 80.1% under the combined audio-visual condition and 67.7% under the visual-only condition.

3.2. The morphology comparison of brainwaves

Figure 2 shows the average brainwaves across all subjects at electrode CZ at the two tone cue stages (i.e., under the visual-only and combined audio-visual conditions). For the brainwaves under the visual-only condition in Fig 2(a), two notable regions are observed (marked with two rectangles). Meanwhile, for the brainwaves under the combined audio-visual condition in Fig 2(b), three notable regions are observed (marked with three rectangles). In Fig. 2, the peak of brainwave of each tone was used as the center, and then a time window of 200 ms was used to calculate the amplitude and latency at electrode CZ.

Multiple paired comparisons with Bonferroni correction were run between the characteristics (i.e., amplitude and latency) of brainwaves across the two tone cuing conditions. The Bonferroni-corrected statistical significance level was set at p<0.01 (α = 0.05). Analysis revealed that for the positive peaks in Fig. 2(a) and Fig. 2(b), both the paired amplitude and paired latency under the audio-visual condition were significantly larger (p<0.01) than those under the visual-only condition. In addition, based on multiple paired comparisons with Bonferroni correction, the amplitudes and latency of four Mandarin tones in five windows were post hoc tested. The Bonferroni-corrected statistical significance level was set at p<0.01 (α = 0.05). At the visual-only condition in Fig. 2(a), for the positive peaks, there was no significant difference between the amplitudes of any two tones (p>0.01), and no significant differences between the latencies of any two tones (p>0.01). The negative peak in Fig. 2(a) had the same results (i.e., no significant different among amplitudes or latencies) as the positive peak in Fig. 2(a). At the combined audio-visual condition, there was no significant difference between the amplitudes of any two tones (p>0.01) in the first (negative) peak in Fig. 2(b). However, for the latencies, there were significant differences between tone 1 and tone 2 (p<0.01), tone 1 and tone 3 (p<0.01), tone 3 and tone 4 (p<0.01), and there was no significant differences between the other paired tones (p>0.01). For the amplitudes of the second (positive) peak in Fig. 2(b), only tone 2 and tone 4 had significant difference (p<0.01), and the latencies between any two tones had significant differences (p<0.01). Finally, for the amplitudes of the third (negative) peak in Fig. 2(b), there were no significant differences (p>0.01) between any two tones. The latency of tone 2 or tone 3 was significantly different (p<0.01) with that of tone 4.

Actual tone \ Predicted tone	Tone-1	Tone-2	Tone-3	Tone-4
Tone-1	**67.3%**	12.7%	10.8%	9.2%
Tone-2	11.5%	**68.4%**	11.1%	9.0%
Tone-3	13.4%	10.4%	**67.3%**	9.0%
Tone-4	9.9%	10.4%	12.0%	**67.7%**

(a)

Actual tone \ Predicted tone	Tone-1	Tone-2	Tone-3	Tone-4
Tone-1	**82.9%**	8.1%	5.8%	3.2%
Tone-2	9.2%	**76.7%**	7.9%	6.2%
Tone-3	3.7%	10.6%	**77.9%**	7.8%
Tone-4	4.8%	5.1%	7.2%	**82.9%**

(b)

Figure 3: *The confusion matrices of Mandarin tone classification under the (a) visual-only and (b) combined audio-visual conditions.*

3.3. The confusion matrices of Mandarin tone classification

The confusion matrices of tone classification under the two tone cuing conditions are shown in Fig. 3. Multiple paired comparisons with Bonferroni correction were run between the accuracies across the two tone cuing conditions. The Bonferroni-corrected statistical significance level was set at p<0.01 (α = 0.05). Analysis revealed that for each imaginary Mandarin tone, the average classification accuracy under the audio-visual condition was significantly larger (p<0.01) than that under the visual-only condition. Furthermore, post hoc tests, according to multiple paired comparisons with Bonferroni correction, were run between the classification accuracies across the 4 imaginary Mandarin tones under each tone cuing condition. The Bonferroni-corrected statistical significance level was set at p<0.01 (α = 0.05). Analysis showed that under the visual-only condition in Fig. 3(a), there was no significant differences (p>0.01) between any paired tone classification accuracies. However, under the audio-visual condition in Fig. 3(b), both the classification accuracy of tone 1 or tone 4 was significantly larger (p<0.01) than that of tone 2 or tone 3. There was no significant difference (p>0.01) between the tone classification accuracies of tone 1 and tone 4, and no significant difference (p>0.01) between the tone classification accuracies of tone 2 and tone 3.

4. Discussion and conclusions

This work carried out a Mandarin tone imagery experiment, and specially examined potential factors affecting the accuracies of classifying 4 Mandarin tones from EEG signals recorded in the Mandarin tone imagery experiment, including the effects of electrode configuration and tone cuing. Table 1 shows that under the conditions of using 2 electrodes from a specific brain region, the accuracies of classifying four tones are low, i.e., from 31.6% to 39.9 % under the visual condition, and from 39.2% to 55.2% under the combined audio-visual condition. Broca's area has the highest average accuracies of tone classification under the two conditions (i.e., 39.9 % and 55.2%), while Wernicke's area has the lowest average accuracies of tone classification (31.6% and 39.2%). This may be attributed to the fact that different brain regions have different language functions, i.e., Broca's area and Wernicke's area are mainly responsible for language expression and language comprehension/memory, respectively, and the activation of primary motor cortex is affected by the vocal organ [24-25]. When all three brain regions (i.e., with 6 electrodes in this work) examined are used for tone classification, the accuracies of tone classification are significantly improved to 67.7% and 80.1% under the two conditions. This indicates that, depending on the tone cuing condition, the Mandarin tone imagery task is a complex process involving several language-related brain regions. Hence, when more language-related brain regions are involved in the EEG-based classification of imaginary Mandarin tones, more important features could be fused from the EEG signals, favoring the Mandarin tone classification performance.

The results of the present work showed that the combined audio-visual condition (i.e., using auditory and visual stimuli at the tone cue stage) yielded better tone classification performance than the visual-only condition (i.e., using visual stimuli at the tone cue stage) (see Table 1 and Figure 3), suggesting the influence of tone cuing on Mandarin tone imagery. Analyses were carried to explore the possible mechanism accounting for the influences of the two tone cuing methods in this work. It was seen in Fig. 2(a) that under the visual-only condition, the amplitudes and latencies of the four brain responses (corresponding to four tones) at the tone cue stage were not significantly different (or not separable), suggesting that visual-only stimuli caused similar levels and patterns of brain activation. To some extent, this was consistent with the tone classification accuracies shown in Fig. 3(a), whereas at the tone imagination stage, similar classification accuracies (i.e., 67.3%, 68.4%, 67.3%, and 67.7%) were achieved for the four Mandarin tones. On the contrary, as shown in Fig. 2(b) under the combined audio-visual condition, the four brain responses (corresponding to four tones) have notably different waveform, including their amplitudes and latencies. These separable waveforms of the four brain responses are beneficial for classify each tone, which partially accounts for the better tone classification accuracies observed in Fig. 3(b) under the combined audio-visual condition. By comparing the EEG signals recorded under the two conditions (i.e., visual-only vs. combined audio-visual in Fig. 2), it is seen that under the audio-visual condition, the brain was more activated with a significantly large amplitude at the tone cue stage, which might also favor tone classification at the tone imagination stage. Further studies are warranted to investigate the relation between the tone cue and imagination stages.

In conclusion, this work studied the Mandarin tone classification with cortical EEG signals in a Mandarin tone imagery task. Specially, this work assessed the effects of electrode configuration and tone cuing on the tone classification accuracy. Results indicated that the involvement of more activated language-related brain regions provided a more accurate classification of imaginary Mandarin tones than that of a specific brain region. The tone cuing condition with audio-visual stimuli yielded better classification accuracies than that with visual-only stimuli. This was possibly because audio-visual stimuli were more effective than visual-only stimuli in promoting the separable activation of brain regions at the tone cue stage.

5. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61971212), the Basic Research Foundation of Shenzhen (Grant No. GJHZ20180928155002157), and High-level University Fund G02236002.

6. References

[1] Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., and Chang, E. F., "Reconstructing speech from human auditory cortex," *PLOS Biology*, 10:e1001251, 2012.

[2] Hong, B., Wang, Y. J., Gao, X. R., and Gao, S. K., "Visual and auditory brain-computer interfaces," *IEEE Transactions on Biomedical Engineering*, 61(5):1436-1447, 2014.

[3] Zhang, B., Quan, C. Q., and Ren, F. J., "Overview of speech synthesis in development and methods," *Journal of Chinese Computer Systems*, 37(1):186-192, 2016.

[4] Sercan-Ömer, A., Mike, C., Adam, C., Gregory, F. D., Andrew G., Yongguo, K., Xian, L., John, M., Andrew, Y. N., Jonathan, R., Shubho, S., and Mohammad, S., "Deep voice: Real-time neural text-to-speech," *in Proc. of 34th International Conference on Machine Learning, Sydney, Australia*, 195-204, 2017.

[5] Herff, C., Schultz, T., "Automatic speech recognition from neural signals: a focused review," *Frontiers in Neuroscience*, 10:429, 2016.

[6] Leuthardt, E.C., Gaona, C., Sharma, M., Szrama, N., Roland, J., Freudenberg, Z., Solis, J., Breshears, J., and Schalk, G., "Using the electrocorticographic speech network to control a brain-computer interface in humans," *Journal of Neural Engineering*, 8(1):1-11, 2011.

[7] Tian, X., Zarate, J. M., and Poeppel, D., "Mental imagery of speech implicates two mechanisms of perceptual reactivation," *Cortex*, 7(1):1-12, 2016.

[8] Matsumoto, M., and Horib, J., "Classification of silent speech using support vector machine and relevance vector machine," *Applied Soft Computing*, 20(1):95-102, 2013.

[9] DaSalla, C. S., Kambara, H., Sato, M., and Koike, Y., "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Network*, 22(9):1334-1339, 2009.

[10] Saha, P., Muhammad, A-M., and Fels, S., "Speak your mind! Towards imagined speech recognition with hierarchical deep learning," *in Proc. of 20th Annual Conference of the International Speech Communication Association (InterSpeech), Graz*, 141-145, 2019.

[11] Repp, B. H., and Lin, H-B., "Integration of segmental and tonal information in speech perception: A cross-linguistic study," *Journal of Phonetics*, 18(4):481-495, 1990.

[12] Chen, F., and Wong, E. Y. W., "Mandarin tone identification with subsegmental cues in single vowels and isolated words," *Speech, Language and Hearing*, 21(3):183-189, 2017.

[13] Wang, W. S. Y., "The Chinese language," *Scientific American*, 228(2):50-60, 1973.

[14] Shi, M. H., Zhou, C. J., Xie, J., Li, S. Z., Hong, Q. Y., Jiang, M., Chao, F., Ren, W. F., Liu, X. Q., Zhou, D. J., Yang, T. Y., "Electroencephalogram-based brain-computer interface for the Chinese spelling system: a survey," *Frontiers of Information Technology & Electronic Engineering*, 19(3):423-436, 2018.

[15] Kyung, H. K., and Ja, H. K., "Analysis of induced gamma-band activity in EEG during visual perception of Korean, English, Chinese words," *Neuroscience Letters*, 403(3):216-21, 2006.

[16] Wang, L., Zhang, X., Zhong, X. F., and Zhang, Y., "Analysis and classification of speech imagery EEG for BCI," *Biomedical Signal Processing and Control*, 8(6):901-908, 2013.

[17] Yang, B., Li, H., Wang, Q., and Zhang, Y., "Subject-based feature extraction by using fisher WPD-CSP in brain computer interfaces," *Computer Methods and Programs in Biomedicine*, 129(1):21-28, 2016.

[18] Thomas, K. P., Cuntai, G., Lau, C. T., Vinod, A. P., and Kai, K. A., "A new discriminative common spatial pattern method for motor imagery brain computer interfaces," *IEEE Transactions on Biomedical Engineering*, 56(11):2730-2733, 2009.

[19] Grosse-Wentrup, M., and Buss, M., "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, 55(8):1991-2000, 2008.

[20] Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G., "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, 8(4):441-446, 2000.

[21] Dornhege, G., Blankertz, B., and Curio, G., "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Transactions on Biomedical Engineering*, 51(6):993-1002, 2004.

[22] Matsumoto, M., and Hori, J., "Classification of silent speech using adaptive collection," *Computational Intelligence in Rehabilitation and Assistive Technologies*, 16-19, 2013.

[23] Naeem, M., Brunner, C., Leeb, R., Graimann, B., and Pfurtscheller, G., "Seperability of four-class motor imagery data using independent components analysis," *Journal of Neural Engineering*, 3(3):208-216, 2006.

[24] Koheia, I., Toshiob, H., Kenichic, S., Kounosukec, T., Hiroshid, K., and Tatsuya, K., "The effect of visual and auditory enhancements on excitability of the primary motor cortex during motor imagery: a pilot study," *International Journal of Rehabilitation Research*, 35(1): 82-84, 2012.

[25] Besle, J., Bertrand, O., and Giard, M-H., "Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex," *Hearing Research*, 258(2):143-151, 2009.