2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

The Effects of Classification Method and Electrode Configuration on EEG-based Silent Speech Classification

Changjie Pan, Ying-Hui Lai, Fei Chen *Senior Member, IEEE*

Abstract— **The effective classification for imagined speech and intended speech is of great help to the development of speech-based brain-computer interfaces (BCIs). This work distinguished imagined speech and intended speech by employing the cortical EEG signals recorded from scalp. EEG signals from eleven subjects were recorded when they produced Mandarin-Chinese monosyllables in imagined speech and intended speech, and EEG features were classified by the common spatial pattern, time-domain, frequency-domain and Riemannian manifold based methods. The classification results indicated that the Riemannian manifold based method yielded the highest classification accuracy of 85.9% among the four classification methods. Moreover, the classification accuracy with the left-only brain electrode configuration was close to that with the whole brain electrode configuration. The findings of this work have potential to extend the output commands of silent speech interfaces.**

I. Introduction

Brain-computer interface (BCI) is a communication system that does not rely on the normal output pathways composed of peripheral nerves and muscles [1]. It directly connects the human or animal brain with devices, and realizes the control of the machines through the acquisition and processing of brain activity signals.

The field of speech-based BCIs has rapidly developed in recent years [e.g., 2-7], as speech-based BCIs have several advantages over the traditional BCIs. Speech is one of the most common and intuitive means of communication for human beings in daily life, and BCI users will feel more natural, comfortable and easier to operate the speech-based BCIs. It can help patients with language disorders or locked-in syndromes to communicate in a more direct and effective way. On the other hand, speech contains massive and abundant information, and decoding different speech can produce a large variety of output commands in BCI systems.

Recently, lots of inspiring studies have been shown in speech-based BCIs, such as speech reconstruction from physiological signals when people listen or speak [2-4], and classification of different imagined speech tasks [5-7]. However, few studies focused on the classification of different categories of speech. Herff et al. utilized fNIRS signals to recognize three different speaking modes: spoken speech, silently uttered speech and imagined speech, while 5 subjects were asked to produce speech of sentences [8]. The accuracies of pairwise classification were between 65% and 80%, and the average classification accuracy for three speaking modes was 61.0%. In Diener et al.'s study, electromyographic (EMG) signals were recorded from 6 electrodes attached to the skin surfaces of the face and throat which were involved in articulation [9]. Three different speaking modes (i.e., audible speech, whispered speech and silent speech) were classified and the average accuracy for ternary classification was 58.4%. In our previous work [10], three speech modalities (i.e., spoken speech, intended speech and imagined speech) were classified by employing cortical EEG signals. The average ternary classification accuracy was 66.0%, while the pairwise classification accuracy ranged from 71.5% to 85.7%. These studies made a preliminary investigation on the classification of different categories of speech, but no study has compared the performances of different classification methods for silent speech based BCIs or investigated the possibility of speech category classification using limited electrodes.

Intended speech and imagined speech are two modalities in Silent Speech Interface (SSI) [11]. Intended speech means that the speaker silently utters with opened mouth and no audible voice is produced, but it involves parts of the articulation. Imagined speech does not involve any actual articulation, hence no actual sound is produced and the speaker just imagines the sound in their mind internally. The classification of these two speech modalities could extend the output commands for speech-based BCIs. In other words, a BCI system could first identify the categories of speech the user conducts (i.e., intended speech or imagined speech), and then classify the content of speech tasks (e.g., /a/, /u/, or /i/). The increase of output commands in silent speech based BCIs will be of great help to those patients who are unable to produce audible speech.

The aim of this work was to assess and compare the effects of different classification methods and electrode configurations on the classification of intended speech and imagined speech from EEG signals. Feature extraction and classification methods in time, frequency and spatial domains were compared to find the method that yielded the highest accuracy. Besides, the effects of different electrode configurations were assessed on the classification performance when only using limited electrodes.

II. Methods

A. Participants

Four female and seven male participants (ranging from 20 to 30 years old) were recruited to participant in this

Changjie Pan is with Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China.

Ying-Hui Lai is with the Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Chinese Taipei.

Fei Chen is with Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China (phone: +86-75588018554; fax: +86-75588018554; e-mail: fchen@sustech.edu.cn).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

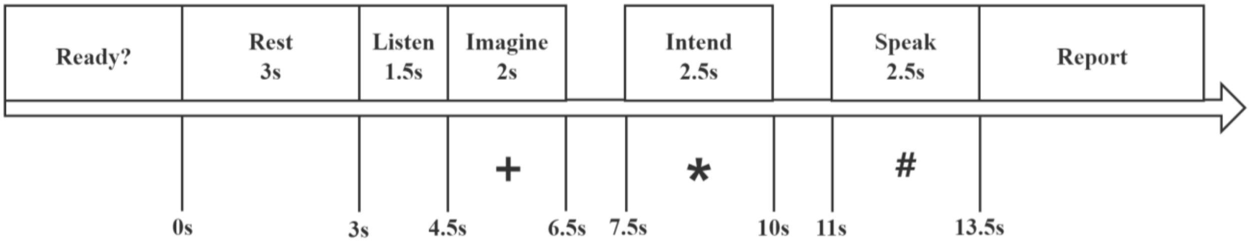

Figure 1. The experimental paradigm. The experiment periods are indicated at the top and corresponding screen displays are shown at the bottom.

experiment. All participants were native Mandarin-Chinese speakers, in good health condition and with no history of neurological or psychological disorders. Informed consents were signed by all participants and the study was reviewed and approved by the Research Ethics Committee of Southern University of Science and Technology.

B. Stimulus and experimental paradigm

The stimuli were 70 Mandarin-Chinese syllables with different vowels, consonants and tones. The experiment consisted of 5 blocks, and each block contained 70 trials in which stimuli were pseudo-randomly selected in different trials. The experimental paradigm is presented in Fig. 1. Participants pressed any key to start the experiment when they were ready. Each trial consisted of 5 successive periods: 1) A 3-second rest period, during which the participants were instructed to look at a blank screen, think nothing and restrain any movements. 2) A 1.5-second listen period, during which the auditory stimulus (recorded at a sampling rate of 16 kHz) pronounced by an adult female native Mandarin-Chinese speaker was presented to participants with an earphone. 3) A 2-second imagined speech period, during which a '+' symbol appeared at the center of the screen till the period ended, and the participant was instructed to stare at the symbol and imagine the pronunciation of the stimulus once without any involvement of vocalization. 4) A 2.5-second intended speech period, during which a '*' symbol appeared, and the participant was instructed to stare at the symbol and silently utter the stimuli with her/his mouth opened and without producing any audible speech. 5) A 2.5-second spoken speech period, during which a '#' symbol appeared, and the participant was instructed to stare at the symbol and speak out the stimulus loudly. There are 1-s gaps between the imagined speech period and the intended speech period, and between the intended speech period and the spoken speech period for period transition. After finishing the five periods of a single trial, the participant could take a short break before pressing the button to go to the next trial. All symbols presented in the screen were in the same font size. The duration of each block was about 20 minutes. The participants were seated comfortably in an acoustically and electrically shielded chamber during the experiment.

C. EEG data recording and pre-processing

The EEG data were recorded with a 64-channel elastic cap (Neuroscience Inc.), which was placed at specific positions following the extended international 10-20 system. The reference electrode was placed at the top of the nose and the ground electrode was attached to the forehead. During the experiment, the impedance between any recording electrode and reference electrode was maintained below 5 kΩ. The recorded EEG data were sampled at 500 Hz. The participants were asked to minimize their movements during the recording in order to avoid possible motion artifacts.

All EEG data were preprocessed with EEGLAB toolbox. The raw data were first re-referenced using the electrodes at contralateral mastoid, and then band-pass filtered between 1 Hz and 30 Hz. Independent Component Analysis (ICA) was applied to remove artifacts during the recording (e.g., electrocardiographic activities, eye blinks, horizontal eye movements, etc.). After the artifact removal, the epochs in the imagined speech periods and the intended speech periods were extracted between 100-ms pre-stimulus and 2000-ms post-stimulus, and then corrected with the baseline of 100-ms pre-stimulus. Considering that the time intervals for imagined or intended speech were generally less than 1000 ms, the EEG data used for classification were finally extracted between 0 ms and 1000 ms. One of the participants was removed from analysis because of the poor data quality.

D. Classification

Four different classification methods were implemented and their performance to classify intended speech and imagined speech was compared.

The first method is the Common Spatial Pattern (CSP) based method, which is widely used and has achieved good performance in motor imagery. The CSP method is a spatial domain feature selection method that uses spatial filters to maximize the discriminability of two classes [12]. After applying the CSP algorithm, the extracted features with a dimension of 16 were trained and classified by a Support Vector Machine (SVM) in this work.

The second method is the time-domain based method adopted by Min et al. [5]. Each epoch was cut into several time segments with 0.2-s length and 0.1-s overlap. Mean value, variance, standard deviation, and skewness of those four features of each channel were extracted within each time segment. Four features in each channel were transformed into a single column feature vector from end to end. Because the dimension of the single column feature vector was too large, a Lasso estimate was applied to reduce the dimension of the feature vector. Then an Extreme Learning Machine (ELM) was used for feature vector classification, and the majority of the predicted labels of time segments within a single epoch was set as the final label of the epoch.

The third method is the frequency-domain based method adopted by Sereshkeh et al. [6]. Discrete wavelet transform (DWT) features were extracted from each epoch using the Daubechies-4 (db4) wavelet. The root-mean-square and

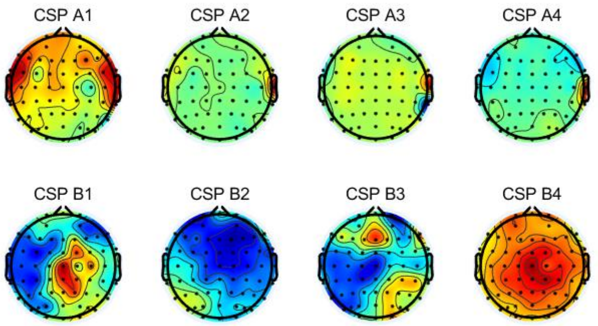

Figure 2. Common spatial patterns of participant s3 for distinguishing imagined speech and intended speech. The top and bottom rows represent the first four patterns for the imagined speech and the intended speech, respectively.

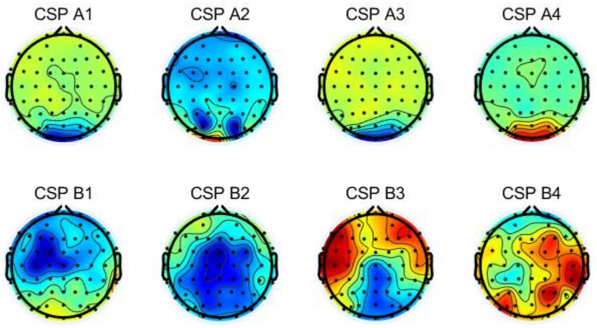

Figure 3. Common spatial patterns of participant s10 for distinguishing imagined speech and intended speech. The top and bottom rows represent the first four patterns for the imagined speech and the intended speech, respectively.

standard deviation of the coefficients from different DWT decomposition level were extracted as features. Then, an artificial neural network, which had one hidden layer with 10 hidden units, was used for classification.

The fourth method is the Riemannian manifold based method [7]. Similar to the CSP based method, features in spatial domain were extracted. The covariance matrix of each epoch was first calculated and the covariance matrices were treated as sample points in the Riemannian space. Then these samples were projected into the Riemannian tangent space and then classified by Linear Discriminant Analysis (LDA).

For each participant, 80% of epochs were randomly chosen as a training set, and the rest 20% were chosen as a testing set for performance evaluation. This procedure was repeated for 20 times and the average of classification accuracies was calculated for each participant.

III. RESULTS

Four methods were compared to search the best method to classify imagined speech and intended speech, and their classification results are presented in Table 1. It is shown that, among the four methods tested, the Riemannian manifold based method achieved the highest average classification accuracy, i.e., 85.9%. The CSP based method yielded a relatively good result, i.e., 81.6%, which was slightly inferior to that of the Riemannian manifold based method. The Sereshkeh's and Min's methods gave lower average classification accuracies of 74.4% and 68.0%, respectively.

TABLE 1. Classification results of intended speech and imagined for four different methods.

Subject ID	Riemannian based method (%)	CSP based method (%)	Min's method (%)	Sereshkeh's method (%)
s1	94.8	90.6	67.2	78.1
s2	93.0	92.8	68.7	78.2
s3	91.2	89.0	77.9	85.9
s4	77.8	64.9	66.7	70.0
s5	91.9	84.8	67.1	74.4
s7	72.5	69.9	57.2	63.0
s8	80.1	75.0	59.7	59.7
s9	70.0	66.2	59.0	59.5
s10	94.1	93.0	82.4	90.8
s11	94.0	89.9	74.7	84.3
Avg.	85.9	81.6	68.0	74.4
Std.	9.7	11.4	8.3	11.1

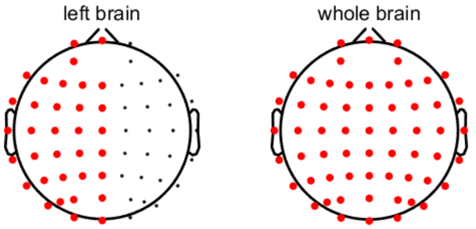

Figure 4. Two adopted electrode configurations. The red dots represent the selected electrodes in the configuration.

TABLE 2. Average classification results across all participants for two different electrode configurations with the Riemannian manifold based method.

	left brain (%)	whole brain (%)
Acc.	82.8	85.9
Std.	10.3	9.8

In order to investigate the effect of utilizing limited electrodes to classify imagined speech and intended speech, common spatial patterns were used to select channels. Figures 2 and 3 show the first four pairs of common spatial patterns of participants s3 and s10, respectively, for distinguishing imagined speech and intended speech. The four patterns of CSP A1-A4 at the top row represent the first four patterns for imagined speech, and CSP B1-B4 at the bottom row represent the first four patterns for intended speech. In Fig. 2, CSP A1-A4 of participant s3 are relatively flat, and no brain area is activated evidently during imagined speech. In contrast, obvious activation is observed in the left hemisphere in CSP B1, B2 and B3 during intended speech. Similar phenomena could also be seen in the topological maps of participant s10 as shown in Fig. 3, where CSP A1-A4 show no obvious activation during imagined speech, while CSP B1 and B3 show distinct patterns in the left hemisphere during intended speech.

The above-mentioned findings suggested that the left hemisphere was more strongly activated in intended speech than in imagined speech and the difference of two speech modalities mainly laid in the left hemisphere. For this reason, the left brain electrode configuration and the whole brain electrode configuration were compared to assess whether effective classification could be achieved with electrodes only located in the left hemisphere. Two adopted electrode configurations are shown in Fig. 4. With the Riemannian manifold based method, the average classification results in the two different electrode configurations are presented in Table 2. It is shown that the classification accuracy with the left-only brain electrode configuration is 82.8%, which is slightly lower than 85.9% with the whole brain electrode configuration.

IV. Discussion and Conclusion

This work distinguished imagined speech and intended speech by employing the cortical EEG signals recorded from scalp. It is shown in Table 1 that the classification performance of the Riemannian manifold based method and the CSP based method outperformed that of the Min's and Sereshkeh's methods, which could be probably attributed to the obvious difference observed between the spatial features of the EEG signals of imagined speech and intended speech. The difference of EEG features between two speech modalities may not be distinctly observed in time domain or frequency domain but in spatial domain. Furthermore, the Riemannian manifold method yielded a slightly better classification accuracy than the CSP based method, indicating the robustness against noise for the Riemannian manifold based method as suggested in early studies [e.g., 13].

Intended speech could be treated a truncated version of spoken speech, because intended speech includes nearly intact articulation but without actual sound production. In [14], fMRI experiments indicated that spoken speech showed greater response than imagined speech in left premotor cortex, left primary left insula, and left superior temporal gyrus. In a word production experiment using ECoG signals [15], it was found that Wernicke's area and Broca's area in the left brain were activated during the production of spoken speech rather than imagined speech. The above-mentioned studies may support that the difference of intended speech and imagined speech is mainly located in the left hemisphere, which dominates the speech production processing. This could also partially explain why the classification accuracy with the electrodes only in the left brain is almost as good as that with all electrodes, as shown in Table 2.

In conclusion, this study investigated the effects of classification methods in different domains and electrode configurations on the classification of intended speech and imagined speech using EEG signals. Four classification methods were compared, and the Riemannian manifold based method yielded the best average classification accuracy among all methods tested, manifesting the importance of spatial-domain features over time- or frequency-domain features. Furthermore, the findings in this work indicated that a good classification accuracy could be achieved when only using the electrodes in the left hemisphere. The effective classification for imagined speech and intended speech has potential to extend the output commands of speech based BCIs, particularly in silent speech interfaces, in the future.

V. Acknowledgements

This work was supported by Shenzhen Sustainable Support Program for High-level University (20200925154002001), and High-level University Fund of Southern University of Science and Technology.

VI. References

[1] T.M. Vaughan, "Guest editorial brain-computer interface technology: a review of the second international meeting," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 94–109, 2003.

[2] H. Akbari, B. Khalighinejad, J.L. Herrero, A.D. Mehta, and N. Mesgarani, "Towards reconstructing intelligible speech from the human auditory cortex," *Scientific Reports*, vol. 9, no. 1, pp. 874, 2019.

[3] C. Herff, L. Diener, M. Angrick, E. Mugler, M.C. Tate, M.A. Goldrick, D.J. Krusienski, M.W. Slutzky, and T. Schultz, "Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices," *Frontiers in Neuroscience*, vol. 13, pp. 1267, 2019.

[4] G.K. Anumanchipalli, J. Chartier, and E.F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, 493–498, 2019.

[5] B. Min, J. Kim, H.J. Park, and B. Lee, "Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram," *BioMed Research International*, vol. 2016, pp. 2618265, 2016.

[6] A.R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Online EEG classification of covert speech for brain-computer interfacing," *International Journal of Neural Systems*, vol. 27, no. 8, pp. 1750033, 2017.

[7] C.H. Nguyen, G.K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features," *Journal of Neural Engineering*, vol. 15, no. 1, pp. 016002, 2018.

[8] C. Herff, F. Putze, D. Heger, C. Guan, and T. Schultz, "Speaking mode recognition from functional Near Infrared Spectroscopy," in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1715–1718, 2012.

[9] L. Diener, S. Amiriparian, C. Botelho, K. Scheck, D. Küster, I. Schuller, B.W. Trancoso, and T. Schultz, "Towards silent paralinguistics: deriving speaking mode and speaker ID from electromyographic signals," in *Proceedings of 21st Annual Conference of the International Speech Communication Association*, 2020.

[10] C. Pan, Z. Liu, and F. Chen, "Speech modality classification with cortical EEG signals," in *Proceedings of International IEEE EMBS Conference on Neural Engineering*, pp. 69–72, 2021.

[11] B. Denby, T. Schultz, K. Honda, T. Hueber, J. Gilbert, and J. Brumberg, "Silent speech interfaces," *Speech Communication Journal*, vol. 52, no. 4, pp. 270–287, 2010.

[12] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.

[13] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, 2017.

[14] L.I. Shuster, and S.K. Lemieux, "An fMRI investigation of covertly and overtly produced mono- and multisyllabic words," *Brain and Language*, vol. 93, no.1, pp. 20–31, 2005.

[15] X. Pei, E.C. Leuthardt, C.M. Gaona, P. Brunner, J.R. Wolpaw, and G. Schalk, "Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition," *Neuroimage*, vol. 54, no.4, pp. 2960–2972, 2011.