33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# An Auditory Brain-Computer Interface Using Virtual Sound Field

Haiyang Gao, Minhui Ouyang, Dan Zhang, Bo Hong* *Member, IEEE*

***Abstract*—Brain-computer interfaces (BCIs) exploring the auditory communication channel might be preferable for amyotrophic lateral sclerosis (ALS) patients with poor sight or with the visual system being occupied for other uses. Spatial attention was proven to be able to modulate the event-related potentials (ERPs); yet up to now, there is no auditory BCI based on virtual sound field. In this study, auditory spatial attention was introduced by using stimuli in a virtual sound field. Subjects attended selectively to the virtual location of the target sound and discriminated its relevant properties. The concurrently recorded ERP components and the users' performance were compared with those of the paradigm where all sounds were presented in the frontal direction. The early ERP components (100-250ms) and the simulated online accuracies indicated that spatial attention indeed added effective discriminative information for BCI classification. The proposed auditory paradigm using virtual sound field may lead to a high-performance and portable BCI system.**

## I. Introduction

Brain-computer interfaces (BCIs) based on the visual modality are effectively and widely used [1-3]. However, most of them are not suitable for patients who have compromised vision or lose the control of their eye movement. In this case, the BCI using auditory modality might be a solution for them.

The auditory P300-based BCI has been explored [4-5] and applied to the amyotrophic lateral sclerosis (ALS) patients [6]. However, users' performance was lower than achieved by a similar visual modality system. Recently, Guo *et al.* used active mental response for building an auditory BCI [7] and the classification accuracy was improved. In their study, the spoken digits of 1-8 were chosen to be the stimulus sounds. The subjects were required to focus on the target digit and make a voluntary response, such as discriminating the laterality property (voice from left or right side) or the gender property (male or female voice). A broad late positive ERP component (LPC) was elicited by introducing such mental response.

Many studies demonstrated that auditory spatial attention could modulate the early ERP components [8-9]. In the hope of improving the performance of the auditory BCI, we also introduced auditory spatial attention. In this study, active mental response proposed by Guo *et al*. was adopted; meanwhile a virtual sound field was used to achieve spatial attention. The subjects attended selectively to the target digit delivered from a certain direction while ignoring the stimulus sounds from other directions. When hearing the target digit, the subjects were required to distinguish its gender property. The N1 and late positive component (LPC) were elicited by target voice and selected as features in target detection, through which the target digit that the subject wants to communicate can be identified. By using the virtual sound field, only a headphone is needed to present the stimulus. Thus, the portability of the BCI system was achieved. In addition, there is potential for online applications for this BCI system. .

This work was supported in part by the National Science Foundation of China under Grant 61071003 and 90820304.

B Hong is with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China (e-mail: hongbo@tsinghua.edu.cn).

H. Gao, M. Ouyang, D. Zhang are with the Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China.

## II. Method

### A. Subjects and Data Recording

Seven subjects (six men, aged 20-25) with normal hearing participated in this experiment. All subjects gave informed consent and were paid for their participation.

The electroencephalogram (EEG) was recorded from 30 scalp electrodes (Electro-Cap, Neuroscan) according to the international 10-20 electrode system, referenced to linked-mastoids. Each electrode's impedance was kept below 5kΩ. Signals were band-pass filtered between 0.05 and 200 Hz with a sampling rate of 1000 Hz.

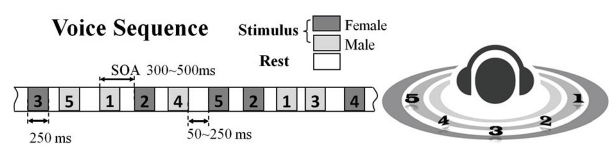


Fig.1. "Spatial paradigm" scheme with voice sequence design: Each digit came from a fixed direction. The angle between two adjacent digits was 45°. The duration of each stimulus was 250ms and the inter-stimulus interval was random in the range of 50ms to 250ms.

### B. Experimental Paradigm and Procedure

The auditory stimulus was a sequence of five spoken Chinese digits (1-5) with the same duration and intensity, each corresponds to a command the subject wants to communicate. In the so called "spatial paradigm" experiment, each digit was presented from a fixed direction, and each digit can be a male or female voice. Fig.1 shows the configuration of the voice sequence.

The virtual sound field was constructed by using the head-related transfer functions (HRTFs) technology [10], which can be simply regarded as the system function of the information channel from the sound source to ears. The HRTFs database and related tools in this study were available at the website of the Institute for Research and Coordination Acoustic/Music (IRCAM, http://www.ircam.fr/accueil.html).

In the present study, the basic stimulus unit is an epoch, in which one of the five digits is presented with a random gender property. One trial consists of five stimulus epochs with different digits. 15 consecutive trials with the same digit as the target digit constitute a block. Each digit was chosen as target digit once in a session experiment. The experiment was carried out in an offline manner, with no feedback to the subjects throughout the experiment.

Subjects were seated in front of a computer screen in a shielded chamber and the air conducted earphones (Etymotic ER2, Etymotic Research, IL) were used to avoid possible electromagnetic interferences from the auditory devices. Before each block, both male and female voices of the target digit were presented to acquaint the subjects with the stimuli. During the experiment, subjects kept their eyes focusing on a cross appearing on the computer screen in order to reduce the eye movement and blinks. In the "spatial paradigm", subject was instructed to pay attention to the direction of the target digit, while ignoring digits from the other directions, then discriminate the target voice's gender and silently speak "male" or "female".

To explore the influence of spatial attention on BCI system performance, we carried out a control experiment called "non-spatial paradigm" by using the active mental response only[7]. In the "non-spatial paradigm", all stimulus voices were delivered from directly ahead (only one spatial location) and the subject should focus on the target digit and discriminate its gender property.

All the subjects participated in two experiment sessions for each paradigm, and the corresponding EEG signals were recorded.

### *C. Data Analysis*

The EEG data of the two paradigms were analyzed using the same procedure. The recorded data were down-sampled to 200Hz and band-pass filtered (1Hz-15Hz). The data epoch was extracted between 0ms and 800ms relative to the stimulus onset, using the pre-stimulus -100~0ms segment as the baseline.

To investigate the modulation effect of spatial attention, the data epochs from the two-session experiments of each subject were group averaged to obtain the grand average responses for the target and non-target respectively. The target and non-target responses in both paradigms were then subjected to statistical analysis.

Electrodes FCz and Pz, where target and non-target stimuli elicited most prominent response differences, were used for an offline classification (Figure 2b). The data epochs (0-800ms) were decimated to 20Hz and concatenated to a 32 dimensional feature vector. As the number of target and non-target epochs were imbalanced (1:4, one digit as target and the rest 4 digits as non-target), traditionally this problem was tackled by randomly discarding a portion of non-target samples [7]. Here we introduced different penalty values for the two types of epochs using a support vector machine (SVM) classifier. The penalty value was set to 1 for non-target epochs and 4 for target epochs. By introducing the penalty value, all the data samples can be fully utilized [11].

To obtain the optimal parameters for the SVM classifier, 5-fold cross-validation was employed. G-mean was employed as the optimization goal for the cross-validation. G-mean is defined as the geometric mean of the positive class accuracy and the negative class accuracy, which is a common metric for the evaluation of classification for imbalanced problems [17].

As the general principle for evoked potentials, the detection of ERP components in the epoch following a single event is not reliable and averaging across trials is always applied for ERP-based BCIs. Here we carried out a simulated online analysis to explore the accuracy of target detection as a function of averaged trial number. For both paradigms, the first session was chosen as the training data set, while the second session was used as the testing data set. In the training data set, vectors from adjacent three trials were averaged, resulting in five target samples and 20 non-target samples per block. The samples were then used to train the classifier in the similar way as mentioned above. In the testing data set, vectors from the first N trials (N=1, ..., 15) in each block were averaged according to the digit. Moving from epoch-based classification to trial-based classification, we used the decision values of each epoch. The decision value represents the distance from the sample to the hyperplane. For example,

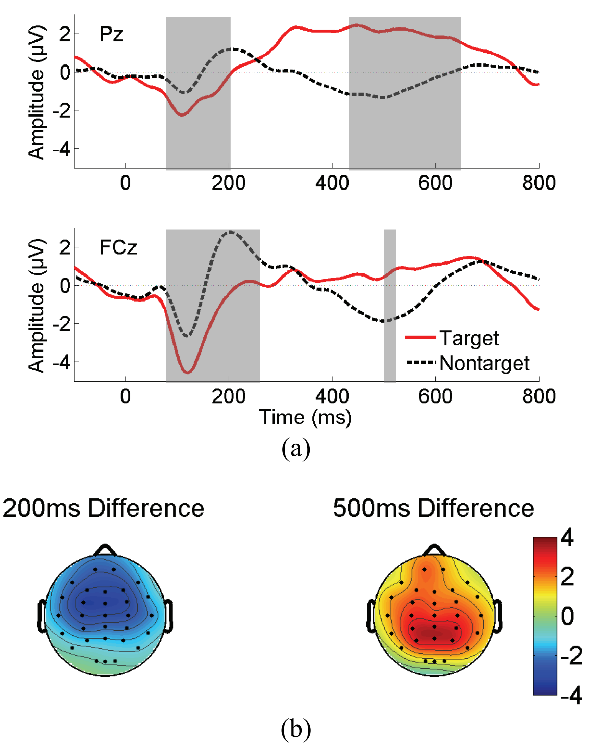


Fig.2. The spatio-temporal patterns of auditory ERP from all subjects' grand average in the "spatial paradigm": (a) Target/nontarget ERPs at Pz and FCz, target/non-target ERPs are plotted with red solid/black dashed line, respectively. Shaded areas in the waveforms show paired t-test results with $p<0.05$. (b) Grand average amplitude difference topographic maps of all subjects at 200ms and 500ms.

a positive value indicated that the sample might be the target sample. The target digit was decided as the digit with the maximum decision value among the 5 decision values of the samples of each block. Finally, the accuracy was calculated for each subject. All analyses were carried out using EEGLAB [12] and LIBSVM [13] toolboxes in MATLAB (Mathworks).

## III. RESULTS

### A. Spatio-Temporal Pattern of Auditory ERP

Fig.2 shows all subjects' grand averaged ERPs and amplitude difference topographic maps in the "spatial paradigm". The target voice elicited an ERP different from that of the non-target. Specifically, the target voice elicited a more evident negative component N1 and the non-target voice evoked an obvious P2 component. The target waveform was significantly different from the non-target waveform during the time periods 100-200ms and 400-600ms. The early difference component was mainly distributed over the frontal and anterior central area, while the late difference component was distributed over the parietal area.

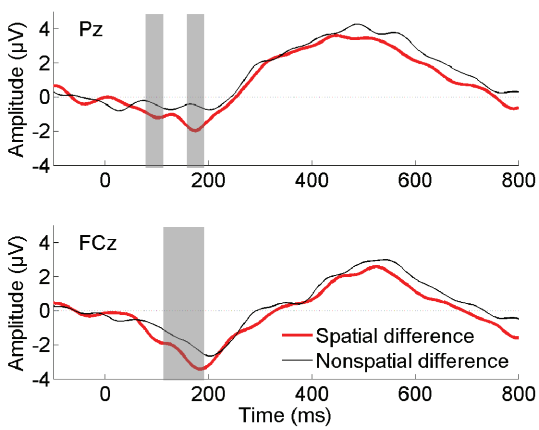


Fig.3. Grand average difference waveforms comparison between two paradigms. Waveforms of "spatial paradigm" are in red bold curves and waveforms of "non-spatial paradigm" are in black normal curves. Shaded areas in the waveforms show the paired t-test results with p<0.05.

### B. Comparison with "Non-Spatial Paradigm"

As shown in Fig.3, the difference waveforms between target and non-target responses of the "spatial paradigm" elicited more negative components during 100-200ms, compared with those of the "non-spatial paradigm".

TABLE I
G-MEAN OF THE TWO PARADIGMS

| Subject | Spatial | Non-spatial |
|---|---|---|
| S1 | 0.6593 | 0.5591 |
| S2 | 0.6724 | 0.6355 |
| S3 | 0.7073 | 0.7054 |
| S4 | 0.6816 | 0.6294 |
| S5 | 0.6318 | 0.6088 |
| S6 | 0.6469 | 0.6182 |
| S7 | 0.8709 | 0.8457 |

The discriminability of target and non-target epochs measured by G-mean values in both paradigms for all individual subjects was shown in Table I. G-mean values were significant larger in spatial paradigm (p=0.0176), revealing a positive benefit of introducing spatial attention.

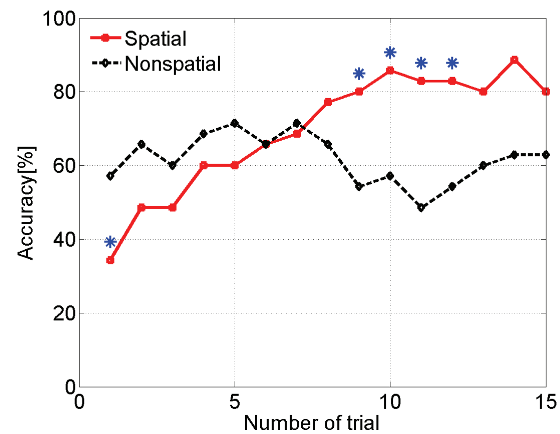


Fig.4 Average target detection accuracy of all subjects as a function of the number of trials averaged. (Blue asterisk indicates a significant difference of accuracy between two conditions.)

### C. Target Detection

The average classification accuracies of all subjects as a function of the number of averaged trials in the two paradigms were shown in Fig.4. As a general trend, the spatial paradigm showed better performance with more averaged trials. The spatial paradigm significantly outperformed the non-spatial paradigm when the number of averaged trial was between 9 and 12

## IV. DISCUSSION AND CONCLUSION

Spatial attention resulted in two prominent ERP components expressing target vs. non-target differences. The LPC component of the target stimulus reflected the subjects' voluntary mental effort in discriminating the stimulus property [7]. The positive peak around 500ms of the target ERPs may be related to the top-down processing of the target stimuli. These two components were the major contributors to the classification.

Auditory stimulus that appears at attended locations elicits larger amplitude ERP components, reflecting basic sensory processing (N1) and higher-level perceptual mechanisms (processing negativity) [14]. Compared with the ERPs of the stimuli from unattended directions, in the "spatial paradigm" the stimulus located at attended direction elicited more negative N1 component. The component around 200ms of the target ERP is negative, different from the obvious P2 component elicited by the unattended stimuli from other directions (Fig. 2). Previous studies suggested that the ERP component (100ms-250ms) which reflects the spatial attention effect will become more distinct when the spatial distance between attended direction and unattended direction gets larger [9]. Our findings on the ERP difference of the "spatial paradigm" and "non-spatial paradigm" (Fig. 3) is consistent with these studies.

The mean accuracies of both paradigms were below 60% when only a single trial was used for detection. The reason is probably that subjects needed some time to adjust themselves from the initial state at the beginning of each block experiment; the lower accuracy of the "spatial paradigm" at the beginning may result in the poor execution of the spatial attention task. However, when 9-12 trials were averaged, the mean accuracy of the "spatial paradigm" was significantly higher than that of the "non-spatial paradigm". With longer training time, the subjects are expected to perform the spatial attention task better and faster, likely resulting in better BCI performance with less trial repetition.

In [15], Schreuder *et al.* tested another auditory BCI paradigm based on spatial attention. The stimuli were presented from five speakers in free-field. The virtual sound field in this study achieved comparable BCI performance. Moreover, it is more portable and convenient for online application. In addition, the semantic stimuli in this study activated more brain regions related to the higher speech function, thus a greater ERP response may be elicited [16]. The subject's task to distinguish the different choices in this study was not demanding, which makes a more user-friendly interface.

The target detection accuracies in the "spatial paradigm" were significantly higher than those of the "non-spatial paradigm". However, the increase in the accuracy was demonstrated using only one testing session data. To get a more convincing result requires additional sessions for testing. In conclusion, the auditory BCI paradigm proposed in this study is a feasible solution for lock-in patients with sight deterioration and healthy people who want to liberate their eyes when using BCI systems.

## V. Acknowledgment

The authors would like to thank Jing Guo for fruitful discussions, and Wei Wu for language correction.

## VI. References

[1] Y. J. Wang, X. R. Gao, B. Hong, C. Jia, and S. K. Gao, "Brain-computer interfaces based on visual evoked potentials - Feasibility of practical system designs," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, 2008.

[2] F. Guo, B. Hong, X. R. Gao, and S. K. Gao, "Brain computer interface using motion-onset visual evoked potential," *Journal of Neural Engineering*, vol. 5, pp. 477–485, 2008.

[3] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R.Wolpaw, "Toward enhanced P300 speller performance," *Journal of Neuroscience Methods*, vol. 167, pp. 15–21, 2008.

[4] E. W. Sellers, A. Kubler, and E. Donchin, "Brain-computer interface research at the university of south Florida cognitive psychophysiology laboratory: the P300 speller". *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 221-224, 2006.

[5] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, et al, "An auditory oddball (P300) spelling system for brain-computer interfaces", *Psychophysiology*, vol. 46, no. 3, pp. 617-625, 2009.

[6] A. Kubler, A. Furdea, S. Halder, E. M. Hammer, and F. Nijboer, "A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-In patients," *Annals of the New York Academy of Sciences*, vol. 1157, *Disorders of consciousness* pp. 90-100, 2006.

[7] J. Guo, S. K. Gao, and B. Hong, "An auditory brain–computer interface using active mental response," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 3, pp. 230-235, 2010.

[8] W. A. Teder-Salejarvi, S. A. Hillyard, B. Roder, H. J. Neville, "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," *Cognitive Brain Research*, vol. 8, no. 3, pp. 213–227, 1999.

[9] R. A. Stephen, A. Claude, "Effects of perceptual context on event-related brain potentials during auditory spatial attention," *Psychophysiology*, vol. 39, no. 2, pp. 625–632, 2002.

[10] P. Hofman, J. V. Opstal, S. Gielen, "Relearning sound localization with new ears," *Nature neuroscience*, vol. 1, no. 5, pp. 417-421, 1998.

[11] Y. C. Tang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics*, vol. 39, no. 1, pp. 281-288, 2009.

[12] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[13] C. Chang and C. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/ libsvm

[14] S. V. Stormer, J. J. Green, J. J. McDonald, "Tracking the voluntary control of auditory spatial attention with event-related brain potentials," *Psychophysiology*, vol. 46, no. 2, pp. 357–366, 2009.

[15] M. Schreuder, B. Blankertz, and M. Tangermann, "A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue," *PLoS ONE*, vol. 5, no.4, 2010.

[16] G. Hickok, D. Poeppel, "Opinion - The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393-402, 2007.

[17] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," in *Proc. 14th International Conference on Machine Learning*, Nashville, 1997, pp. 179–186.