Auditory Brain–Computer Interface Paradigm with Head Related Impulse Response–based Spatial Cues

Chisaki Nakaizumi[1], Koichi Mori[2], Toshie Matsui[1], Shoji Makino[1], and Tomasz M. Rutkowski[1,3,*]

[1]*Life Science Center of TARA, University of Tsukuba, Tsukuba, Japan*
[2]*Research Institute of National Rehabilitation Center for Persons with Disabilities, Tokorozawa, Japan*
[3]*RIKEN Brain Science Institute, Wako-shi, Japan*
Email: tomek@tara.tsukuba.ac.jp
http://bci-lab.info/

Abstract—The aim of this study is to provide a comprehensive test of head related impulse response (HRIR) for an auditory spatial speller brain–computer interface (BCI) paradigm. The study is conducted with six users in an experimental set up based on five Japanese hiragana vowels. Auditory evoked potentials resulted with encouragingly good and stable "aha–" or P300–responses in real–world online BCI experiments. Our case study indicated that the auditory HRIR spatial sound reproduction paradigm could be a viable alternative to the established multi–loudspeaker surround sound BCI–speller applications, as far as healthy pilot study users are concerned.

Keywords-Auditory BCI, P300, EEG, brain signal processing.

I. INTRODUCTION

The majority of successful BCI applications rely on mental visual and motor imagery paradigms, which require long–term user training and good eyesight from the user [1]. Alternative solutions have been proposed recently to utilize spatial auditory [2] or somatosensory modalities [3]–[7]. We propose to extend the previously published spatial auditory BCI (saBCI) paradigm [8] by making use of a head–related–impulse–response (HRIR) [9] for virtual sound images spatialization with headphones–based sound reproduction. We test the concept with six BCI users, in a simple five Japanese hiragana (a, i, u, e, o) spatial speller task as previously proposed, and compared with auditory modality, using a vector–based–amplitude–panning (VBAP) approach [8]. HRIR allows for more precise and fully spatial virtual sound images positioning utilizing even not own user's HRIR measurements [10]. For the experiments reported in the paper we use a public domain CIPIC HRTF Database provided by the University of California, Davis (UC Davis) [9].

The remainder of the paper is organized as follows. In the next section, the experimental setup and the HRIR–based saBCI paradigm are described, together with EEG signal acquisition, pre–processing and classification steps. Next, an analysis of evoked response potentials (ERP) and especially the so–called "aha–' or $P300$ response latencies are described. Finally, classification and discussion of the HRIR–based saBCI paradigm information transfer rate (ITR) results conclude the paper, together with future research directions.

*Corresponding author.

II. MATERIALS AND METHODS

All of the experiments were performed in the Life Science Center of TARA, University of Tsukuba, Japan. Six volunteer users participated in the experiments. The average age of the users was 25 years old (standard deviation 8.06 years old,; one females and five males). The psychophysical and online EEG BCI experiments were conducted in accordance with *The World Medical Association Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects*. The experimental procedures were approved and designed in agreement with the ethical committee guidelines of the Faculty of Engineering, Information and Systems at University of Tsukuba, Japan. The experiments were designed to reproduce the previously reported VBAP–based spatial auditory experiments [8] this time with the more precise HRIR–based spatial auditory BCI stimulus reproduction to simplify the previously reported real sound sources generated with surround sound loudspeakers [2], [11].

A. Spatial Stimulus Sound Generation with HRIR and Reproduction to the User via Headphones

The sound stimuli chosen for the subsequent psychophysical and saBCI experiment were the Japanese synthetic female voice utterances generated using a *say* line–command of the MacOS X computer operating system by Apple Inc. The synthetic female voice labeled as *Kyoko* was chosen. The five Japanese vowels *a, i, u, e,* and *o* were selected. Each stimulus was in a length range of $85 - 120$ ms. The generated synthetic sounds frequency spectra were within a limit of $0 - 11,000$ Hz.

Next the the monaural stimulus sounds were spatialized by using the public domain CIPIC HRTF DATABASE [9] and MATLAB based signal processing functions. The HRIR model measured for the user #124 of the CIPIC HRTF DATABASE was chosen. Each Japanese vowel was set on horizontal plane with the same distance from the user's head at the azimuth locations of $-80°, -40°, 0°,$ $40°, 80°$ for the *a, i, u, e, o* vowels respectively. In order to generate a stereo sound of a vowel placed at a spatial location at azimuth θ and elevation of ϕ the following procedure was applied. Let $h_{l,\theta,\phi}$ and $h_{r,\theta,\phi}$ be the minimum–phase impulse responses from the CIPIC HRTF DATABASE measured at the chosen azimuth θ and

the elevation ϕ at the left (l) and right (r) ears. The respective magnitude responses $H(f)$ obtained from Fourier transformed HRIR, the so–called head related transfer functions (HRTF), could be obtained as $|H_{l,\theta,\phi}(f)|$ and $|H_{r,\theta,\phi}(f)|$. The stereo spatial sound delivered via headphones to the left and right ears respectively could be constructed, in time domain using HRIR, as a two–dimensional signal composed of the left $x_l(t)$ and and right $x_r(t)$ headphone channels as follows,

$$x_l(t) = \sum_{\tau=0}^{n-1} h_{l,\theta,\phi}(\tau)s(t-\tau), \tag{1}$$

$$x_r(t) = \sum_{\tau=0}^{n-1} h_{r,\theta,\phi}(\tau)s(t-\tau), \tag{2}$$

where τ denotes sample time delay and n is the HRIR length as obtained from the CIPIC HRTF DATABASE [9]. The so optioned spatial acoustic stimuli were delivered to the left and right ears of the user using the ear–fitting portable headphones.

B. Psychophysical Experiment Protocol

The psychophysical experiments were conducted to investigate the response time and recognition accuracy by each user for the spatial Japanese vowel stimuli. The original monaural sound stimuli were spatialized using the HRIR–database filters as in equations (1) and (2) using a MAX 6 [12] patcher developed by our team. An user instruction visual interface is depicted in Figure 1. The Japanese hiragana character marked in red color, in the above figure, represented a current spatial *target* which user should attend and respond to using a computer keyboard as soon as possible after the perceived and recognized utterance. The other marked in black Japanese vowel stimuli, indicating *non–targets*, were ignored as in a classical oddball paradigm [1]. The instruction which spatial *target* stimulus to attend was delivered audio–visually, while during the following experimental oddball sequence only auditory modality was used.

The user was instructed in each trial to pay attention to the target stimulus vowel type and its spatial location, as well to respond by pressing a computer keyboard key as soon as the target stimulus was heard. The behavioral response times were collected with the same MAX 6 patcher program used to display instructions (see Figure 1) and to generate the HRIR spatial sounds.

Each single trial was comprised of randomly presented single *target* and four *non–target* vowel stimuli. An inter–stimulus–interval (ISI) was set for a one second in order to give the user time for coronet behavioral response. Each trial with presented five stimuli was separated with two seconds break. We conducted 20 trials for each vowel as the *target* in the sequence. Finally we recorded responses of 100 *target* and 900 *non-target* stimuli in the psychophysical experiment. Results of the psychophysical experiments are discussed in the Section III-A.

Figure 1. The user instruction screen presented on a computer screen during the psychophysical and EEG experiments. The hiragana alphabet characters, displayed on the right in the figure, indicate the five Japanese vowels *a, i, u, e, o,* used in the experiments. The target stimulus in each trial (hiragana character *"e"* in this particular instruction case in the figure) is emphasized by a red color. The interface buttons and experiment setup numbers depicted in the left and top areas of the user interface screen were used by the experimenter to setup and start each trial.

C. EEG Experiment Protocol

In the EEG saBCI online experiment the brainwave signals were collected by a bio–signal amplifier system g.USBamp by g.tec Medical Engineering GmbH, Austria. The EEG signals were captured by sixteen active gel–based electrodes attached to the following head locations *Cz, Pz, P3, P4, Cp5, Cp6, P1, P2, Poz, C1, C2, FC1, FC2,* and *FCz* as in the extended $10/10$ international system [13]. The ground electrode was attached on the forehead at the *FPz* location, and the reference on the left user's earlobe respectively.

The spatial auditory stimuli were delivered, as in the preceding psychophysical experiments, using small headphones. The stimuli were generated using the same MAX 6 patcher as in the psychophysical experiment (see Figure 1) yet in this case triggers were produced by AN BCI2000 [14] EEG acquisition and processing environment used on our study. The triggers were communicated from THE BCI2000 to the THE MAX 6 patcher using an internet user datagram protocol (UDP).

THE BCI2000 environment's visual user interface was used to display the *target* stimulus instructions. The online classification results were also presented using the same display. In a single Japanese hiragana vowel spelling session 10 *target* and 40 *non–target* stimuli were presented. Through the whole five vowels spelling session 50 target and 200 *non–target* stimuli were delivered. The ISI in the EEG experiment was shorter comparing to the preceding psychophysical experiments and set in a range of $300 \sim 330$ ms (allowing for a random jitter within 30 ms to break presentation's unnecessary rhythm), since the brain event related potentials (ERP), especially the "aha–" or P300–responses, had no jitter comparing to the behavioral keyboard presses (compare results in Figures 3 and 4). Each spatial auditory stimulus duration was set to 250 ms. The EEG sampling rate was set to 512 Hz and the notch

filter to remove electric power lines interface of 50 Hz was applied in a rejection band of $48 - 52$ Hz, according to the East Japan power stations specification. The high–pass filter was set at 0.1 Hz and low–pass at 60 Hz cutoff frequencies. Please note that the aim of this research is not noise and artifacts reduction [15], but comprehensive test of HRIR usability for the novel saBCI paradigm.

The EEG signals were processed online by the BCI2000 application which used a stepwise linear discriminant analysis (SWLDA) classifier with features drawn from the $0 - 800$ ms ERP interval [16]. Results of the EEG saBCI paradigm experiments are discussed in the Section III-B.

III. RESULTS

This section presents and discusses results obtained from the psychophysical and EEG saBCI experiments conducted with six subjects as described in the previous sections.

A. Psychophysical Experiment Results

The psychophysical experiment user best response accuracy results are summarized in Table I. All of the participants managed to reach perfect score at least once with 100% accuracy which indicated the first confirmation of the HRIR–based spatial sound usability for the saBCI paradigm. Averaged responses are indicated in a form of confusion matrix depicted in Figure 2. The confusion matrix has visualized on a diagonal the correct response accuracies (the majorities in our case) while on off–diagonals the mistakes. The averaged *targets* were scored above 91% accuracy.

Figure 3 presents analysis of response time delays in form of *violin–plots* depicting distributions, median and interquartile ranges. *Pairwise Wilcoxon rank sum tests* revealed no significant differences among medians of the response distributions further confirming the HRIR–based spatial stimuli for the saBCI experiments.

B. EEG Experiment Results

The results of the EEG experiment are summarized in Figures 4 and 5, as well as in Tables II and III. Table II presents the online classification accuracies of the P300 responses as obtained with the SWLDA classifier. All the six users scored well above the chance level of 20%. There were two users who resulted with 100% accuracies in online saBCI experiments.
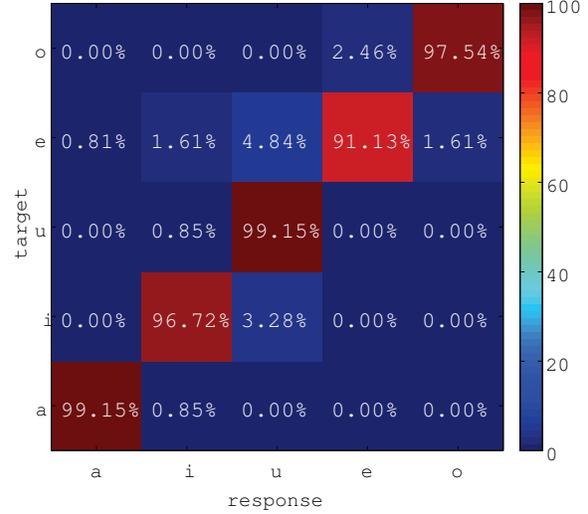

Figure 2. User behavioral response accuracy in form a confusion matrix from the psychophysical experiment depicting averaged results. The numbers indicate the averaged accuracies which have been additionally depicted with color coding. The vertical axis indicates instructed *targets* while the horizontal one the user responses.
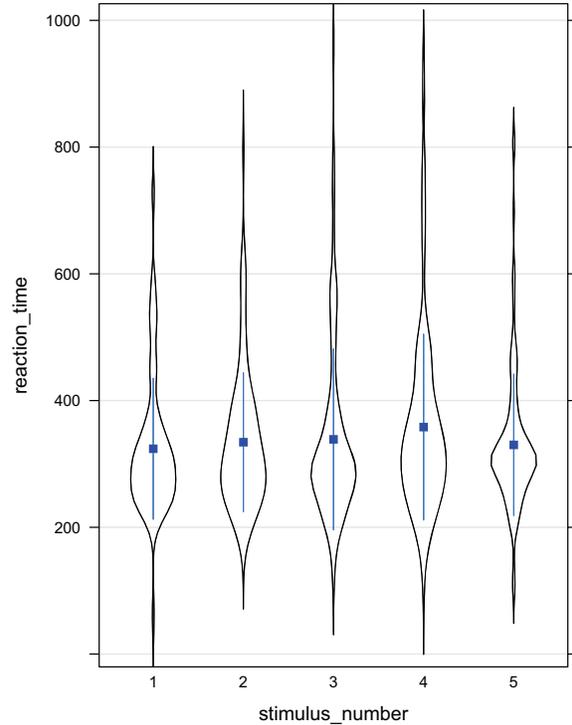

Figure 3. The psychophysical experiment response time probability distributions of all subjects together summarized in form of *the violin–plots* with medians and interquartile ranges depicted. Each number at the horizontal axis represents the stimulus pattern $1 \sim 5$ representing the Japanese hiragana vowels *a, i, u, e, o*, respectively. The horizontal axis represents the reaction (behavioral response) time delays in milliseconds.

Table I
PSYCHOPHYSICAL EXPERIMENT RESULTS (NOTE, THIS IS NOT A BINARY ACCURACY CASE YET THE ONE WITH A THEORETICAL CHANCE LEVEL OF 20%) IN SPATIAL AUDITORY SPELLING TASK

User number	The best psychophysical accuracy
#1	100%
#2	100%
#3	100%
#4	100%
#5	100%
#6	100%
Average:	**100%**

Figure 4 presents the averaged ERP brain responses for all electrodes separately together with standard error bars to *target* (red lines) and *non–target* (blue lines) stimuli. The obvious P300 responses could be clearly seen in the $400 \sim 700$ ms latencies.

Figure 5 summarizes the results as a scalp topography in the top panel with maximum areas under the curve (AUC) [11] values for *target* vs. *non–target* latencies. The topographic plot also reflects the EEG electrodes positions used in the experiments. Although most of the electrodes produced significant P300 potential differences (see also Figure 4), the parietal cortex (the back of the head) resulted with more pronounced differences. The second panel from the top in the Figure 5 indicates the averaged ERP responses to the *target*, while the third panel to the *non–target* stimuli respectively. The bottom panel indicates the AUC of *target* versus *non-target* responses clearly confirming the $400 \sim 700$ ms latencies usability for the subsequent classification.

The presented EEG ERP results confirmed that the HRIR–based spatial auditory stimuli evoked the P300 response possible to classify in the online saBCI experiments with 10–trails averaging.

Table III reports the obtained ITR scores in a range of $4.80 \sim 9.29$ bit/min. The average ITR for the six participants in our study was 6.29 bit/min. The ITR was calculated as follows [11]:

$$ITR = V \cdot R \qquad (3)$$
$$R = \log_2 N + P \cdot \log_2 P + \qquad (4)$$
$$+ \quad (1 - P) \cdot \log_2 \left(\frac{1 - P}{N - 1} \right)$$

where V was the classification speed in selections/minute (4 selections/minute for this study); R represented the number of bits/selection; N was the number of classes (5 in this study); P was the classifier accuracy as in the Table II.

IV. CONCLUSIONS

We conducted a series of psychophysical and online EEG saBCI paradigm experiments in order to evaluate the HRIR–based spatial sound generation system usability. In the psychophysical experiment, the HRIR–based spatial stimuli resulted with close to perfect accuracies. The
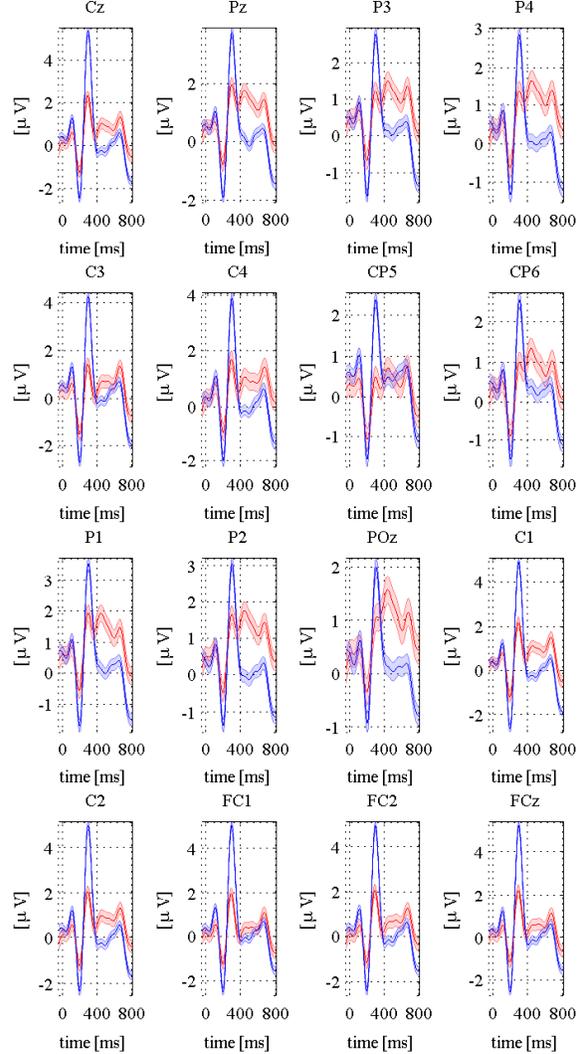

Figure 4. Grand mean averaged EEG ERP together with standard error bars for all participating users in the study plotted for each electrode separately. Red lines *targets* and blue *non–targets*.

online saBCI EEG experiment results also show that the HRIR–based spatial sound reproduction method allows for even 100% accuracy as shown with two subject results. This result suggests a great potential in simple headphone–based HRIR spatial sound generation for the saBCI. The

Table II

TEN AVERAGED TRIALS BASED BCI ACCURACY (NOTE, THIS IS NOT BINARY P300 CLASSIFICATION RESULT BUT RESULTING SPELLING RESULT WITH A THEORETICAL CHANCE LEVEL OF 20%) IN SPATIAL AUDITORY BCI SPELLING TASK USING THE CLASSICAL SWLDA CLASSIFIER

User number	Online BCI experiment SWLDA accuracy
#1	80%
#2	80%
#3	100%
#4	100%
#5	80%
#6	80%
Average:	**86.67**%

Table III

TEN AVERAGED TRIALS BASED SPELLING ACCURACY (SEE TABLE II) ITR RESULTS

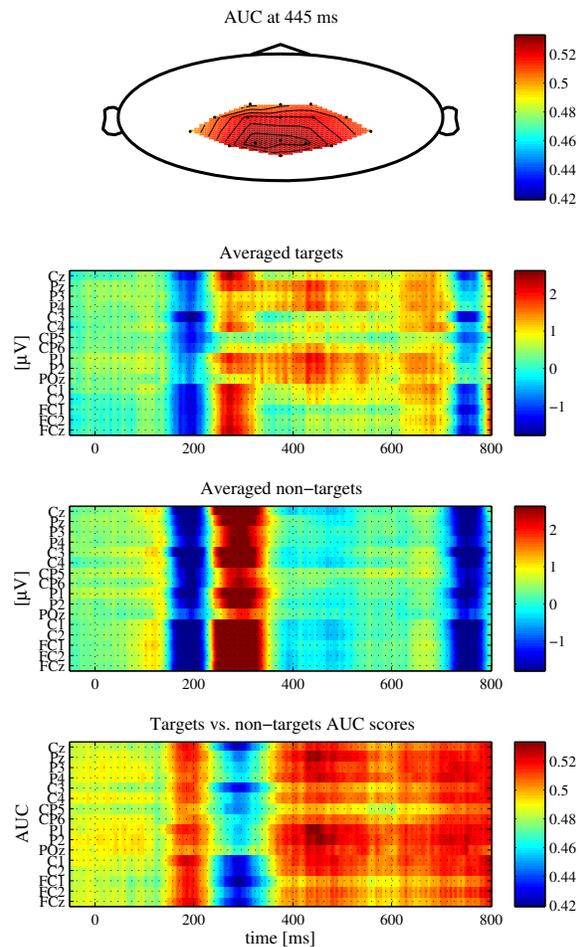| User number	ITR scores
#1	4.80 bit/min
#2	4.80 bit/min
#3	9.29 bit/min
#4	9.29 bit/min
#5	4.80 bit/min
#6	4.80 bit/min
Average:	**6.29** bit/min

Figure 5. Grand mean averaged ERP and AUC scores leading to final classification results of the all participating users. The top panel presents the head topography at the maximum AUC score obtained from the bottom panel. The second panel presents averaged ERP responses to the target, while the third to *non–target* stimuli respectively. The bottom panel visualizes the AUC analysis results of *target* versus *non–target* response distribution differences.

presented results confirmed the hypothesis of the possibility to simplify spatial sound–based saBCI paradigms.

This saBCI scored with high accuracy results without bulky multiple spatially distributed loudspeakers to generate stimuli. Nevertheless, current pilot study is not yet ready to compete with faster visual BCI spellers, for example. Furthermore, it is necessary to improve ITR for comfortable spelling. We plan to continue research with larger number of sound stimuli, with shorter ISI, and with more complex spatial sound patterns.

AUTHOR CONTRIBUTIONS

Designed and performed the EEG experiments: CN, TMR. Analyzed the data: CN, TMR. Conceived the concept of the HRIR–based saBCI paradigm: TMR, KM. Supported the project: SM, KM, TM. Wrote the paper: CN, TMR.

REFERENCES

[1] J. Wolpaw and E. W. Wolpaw, Eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012.

[2] T. M. Rutkowski, A. Cichocki, and D. P. Mandic, "Spatial auditory paradigms for brain computer/machine interfacing," in *International Workshop On The Principles and Applications of Spatial Hearing 2009 (IWPASH 2009) - Proceedings of the International Workshop*, Miyagi-Zao Royal Hotel, Sendai, Japan, November 11-13, 2009, p. P5.

[3] G. Muller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces?" *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 1, pp. 30–37, March 2006.

[4] A.-M. Brouwer and J. B. F. Van Erp, "A tactile P300 brain-computer interface," *Frontiers in Neuroscience*, vol. 4, no. 19, 2010. [Online]. Available: http://www.frontiersin.org/neuroprosthetics/10.3389/fnins.2010.00019/abstract

[5] M. van der Waal, M. Severens, J. Geuze, and P. Desain, "Introducing the tactile speller: an ERP-based brain–computer interface for communication," *Journal of Neural Engineering*, vol. 9, no. 4, p. 045002, 2012. [Online]. Available: http://stacks.iop.org/1741-2552/9/i=4/a=045002

[6] H. Mori, Y. Matsumoto, Z. R. Struzik, K. Mori, S. Makino, D. Mandic, and T. M. Rutkowski, "Multi-command tactile and auditory brain computer interface based on head position stimulation," in *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*. Asilomar Conference Center, Pacific Grove, CA USA: Graz University of Technology Publishing House, Austria, June 3-7, 2013, p. Article ID: 095.

[7] T. Kaufmann, E. M. Holz, and A. Kuebler, "Comparison of tactile, auditory and visual modality for brain-computer interface use: A case study with a patient in the locked-in state," *Frontiers in Neuroscience*, vol. 7, no. 129, 2013. [Online]. Available: http://www.frontiersin.org/neuroprosthetics/10.3389/fnins.2013.00129/abstract

[8] M. Chang, N. Nishikawa, Z. R. Struzik, K. Mori, S. Makino, D. Mandic, and T. M. Rutkowski, "Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms," in *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*. Asilomar Conference Center, Pacific Grove, CA USA: Graz University of Technology Publishing House, Austria, June 3-7, 2013, p. Article ID: 156.

[9] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*. IEEE, 2001, pp. 99–102.

[10] J. Schnupp, I. Nelken, and A. King, *Auditory Neuroscience - Making Sense of Sound*. MIT Press, 2010.

[11] 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, p. e9813, 04 2010. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0009813

[12] "Max 6," 2012. [Online]. Available: http://cycling74.com/

[13] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, no. 4, pp. 1600 – 1611, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811906009724

[14] G. Schalk and J. Mellinger, *A Practical Guide to Brain–Computer Interfacing with BCI2000*. Springer-Verlag London Limited, 2010. [Online]. Available: http://www.schalklab.org/research/bci2000

[15] T. M. Rutkowski, D. P. Mandic, A. Cichocki, and A. W. Przybyszewski, "EMD approach to multichannel EEG data - the amplitude and phase components clustering analysis," *Journal of Circuits, Systems, and Computers (JCSC)*, vol. 19, no. 1, pp. 215–229, 2010. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218126610006037

[16] Z. Cai, S. Makino, and T. M. Rutkowski, "Brain evoked potential latencies optimization for spatial auditory brain-computer interface," *Cognitive Computation*, p. (online first), August 2013. [Online]. Available: http://link.springer.com/article/10.1007%2Fs12559-013-9228-x