Variable Sound Elevation Features for Head–related Impulse Response Spatial Auditory BCI

Chisaki Nakaizumi*, Shoji Makino*, and Tomasz M. Rutkowski*,†,▽
*Department of Computer Science and Life Science Center of TARA, University of Tsukuba
Tsukuba, Ibaraki, Japan
E-mail: tomek@bci-lab.info (▽corresponding author)
†RIKEN Brain Science Institute
Wako-shi, Saitama, Japan

***Abstract*—This paper presents a study of classification and EEG feature improvement for a spatial auditory brain–computer interface (saBCI). This study provides a comprehensive test of a head–related impulse response (HRIR) cues for the saBCI speller paradigm. We present a comparison with previously developed HRIR–based spatial auditory modalities. We propose and optimize the three types of sound spatialization settings using a variable elevation in order to evaluate the HRIR efficacy for the saBCI. Three experienced and seven naive BCI users participate in the three experimental setups based on ten presented Japanese syllables. The obtained EEG auditory evoked potentials (AEPs) result with encouragingly good and stable P300 responses in online saBCI experiments. We analyze the differences and dispersions of saBCI command accuracies, as well as the individual user accuracies for various spatial sound locations. Our case study indicates that the participating users could perceive elevation in the saBCI experiments using the HRIR measured from a general head model.**

I. Introduction

BCI is a technology that uses brain neuronal signals to operate a computer without any muscle movements [1]. Therefore, it is expected to provide a speller for disabled people such as patients suffering from the amyotrophic lateral sclerosis (ALS) [2], [3], a coma or a locked–in syndrome (LIS) [4]. Although, a contemporary and successful visual BCI modality could provide a fast speller [5], the advanced patients who are in the LIS state cannot use it because they lose any intentional muscle control including even eye blinks [1], [3], [4]. Auditory BCI can be an alternative method because it does not require a good sight or eye movements [2], [5]–[10].

We propose an alternative solution, which extends the previously developed by our group spatial auditory BCI (saBCI) paradigms [5], [6] by making use of a head related impulse response (HRIR) for the virtual spatial sound images reproduction with headphones. Our research target is the virtual sound saBCI using the HRIR–based spatial cues to create the non–invasive and auditory stimulus–driven paradigm, which does not require a long training. HRIR appends interaural–intensity–differences (IID), interaural–time–differences (ITD), and spectral modifications to create the spatial stimuli, while a vector–based amplitude panning (VBAP) appends only the IID [9]. The HRIR allows for more precise and fully spatial virtual sound images positioning utilizing even the user not own HRIR measurements [11].

In our previous study [12] we evaluated saBCI feasibility with the HRIR–based spatial sound generator, and compared it with a formerly reported vector–based–amplitude–panning (VBAP)–based spatial auditory experiments [5]. The above study used only five Japanese vowels distributed horizontally every 40° in front of a subject head. The pilot study resulted with clear P300 responses and it has shown that HRIR modality could improve spelling accuracy comparing to the conventional spatial sound generation methods.

In the study presented in this paper we introduce ten Japanese kana syllables–based speller using the sound elevation features as the second step. We also compare various sound elevation settings. However, it is usually difficult to precisely perceive various sound elevations using not the user own HRIR, because the elevation perception is highly individual and it is influenced by the shape of an auricle [11]. Unfortunately, it would be very difficult to record a bedridden ALS–patient own HRIR. Therefore, we propose to test the HRIR–based saBCI efficacy using a general head model (KEMAR). We conduct and EEG experiments in online saBCI ten Japanese kana syllables spelling experiments in order to test our research hypothesis of the KEMAR HRIR–based paradigm practical feasibility.

From now on the paper is organized as follows. The next section describes the experimental set up including the HRIR–based spatial sound design together with EEG signal acquisition, pre–processing and classification steps. The third section focuses on an analysis of event related potentials (ERPs) and especially the $P300$ response latencies leading to saBCI speller classification and information transfer rate (ITR) results in comparison to the conventional method. Finally, conclusions and future research directions are drawn.

II. Materials and Methods

In order to prepare the saBCI stimuli to create an oddball paradigm generating the P300 responses [1], all spatial sound images were created using HRIR of the general head model. Ten Japanese syllables with a vowel "a" were selected for as sound stimuli in this project. The spoken syllables were taken from a public Japanese sound dataset of a female speech [13].

The monaural sound stimuli were spatially distributed using a public domain KEMAR'S HRTF DATABASE provided by the MIT Media Lab [14]. In order to generate a stereo sound placed at a spatial location at an azimuth of θ and an 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 KEMAR HRTF DATABASE measured at the chosen azimuth θ and the elevation ϕ at the left (l) and right (r) ears. 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 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 HRTF DATABASE [15].

The so created spatial acoustic stimuli were delivered to the left and right ears of the user through the ear–fitting portable headphones SENNHEISER CX 400II. Three proposed spatial settings to evaluate the feasibility of the sound elevation using the HRIR for the saBCI are presented in Figure 1.

The first spatial stimulus setting included only the horizontal sound images' placement of the ten saBCI commands. The sound images were distributed at five directions every 45° on the horizontal plane at the user ears' level. Two types of commands were delivered from the same spatial directions.

The second stimulus setting included the sound elevation variability. Ten stimuli were distributed among the all different directions. Five commands were localized at elevation of 50° from the user ears' level and every 45° horizontally.

The third setting included also the variable elevation and an additional option of a frequency modulation appended to discriminate the sound images originating from various elevations.

In the psychoacoustics there is a well–known property called a tonal bell causing a higher–frequency sound to be perceived as originating from a higher elevation [11]. To simulate the above effect we shifted in frequency domain the stimuli, which shall be perceived at higher elevations, using the TANDEM-STRAIGHT method [16]. The sound stimuli to be perceived at elevation of 50° had shifted up their fundamental frequencies by 24 Hz.

All of the experiments reported in this paper were performed in the Life Science Center of TARA, University of Tsukuba, Japan. Ten healthy users participated in our study. They were comprised of seven naive and three experienced BCI users. The average age was of 24.7 years old (standard deviation 6.48 years old, three males and seven females). The online EEG saBCI 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

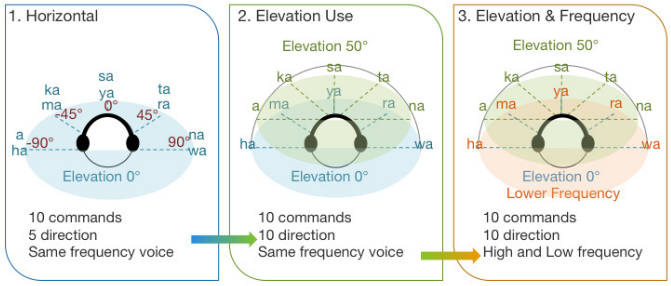

Fig. 1. A diagram of the three proposed virtual spatial stimulus sound settings. The left box depicts the horizontal only placement–based setting; the middle the elevation use–based case; and the right box the elevation use with an additional frequency shifting set up.

of Engineering, Information and Systems at University of Tsukuba, Japan. The users participated voluntarily without any monetary compensation. The users signed informed consent before taking part in the study.

The online EEG experiments were conducted to investigate whether P300 responses could be evoked in the various spatial sound settings and to compare the saBCI classification accuracies, as well as an efficacy of each set up. The brain 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 g.LADYbird 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 [1]. The ground electrode was attached on the forehead at the *FPz* location, and the reference on the user's left earlobe respectively. An in–house extended BCI2000 [17] software was used for the saBCI experiments to present stimuli and display online classification results.

A single EEG experimental session was comprised of one training and two test runs for each spatial stimulus setting. In the training run, EEG brainwaves were recorded and next classifier parameters were calculated. In order to spell a single character, 15 *targets* and 135 *non-targets* were presented for ERP response averaging. In the test runs, user spelled two sets of the three words that the covered ten *targets* (*a–ta–ma; ha–ra; sa–wa–ya–ka–na; wa–ta; ha–na–ya; a–ka–ra–sa–ma*). For spelling single character, 5 *target* and 45 *non-target* stimuli were presented continuously as for single syllable (five ERPs averaging scenario). Experiment order of every spatial sound setting and spelled words were randomized for each subject. The sound stimulus duration was set to 300 ms and the inter–stimulus–interval (ISI) to 150 ms. The EEG sampling rate was of 512 Hz and a notch filter to remove electric power lines interface of 50 Hz was applied in a rejection band of $48 \sim 52$ Hz. The band–pass filter was set at 0.1 Hz and 60 Hz cutoff frequencies.

The acquired EEG brain signals were classified online by the in–house extended BCI2000 application using a stepwise linear discriminant analysis (SWLDA) [18] classifier with

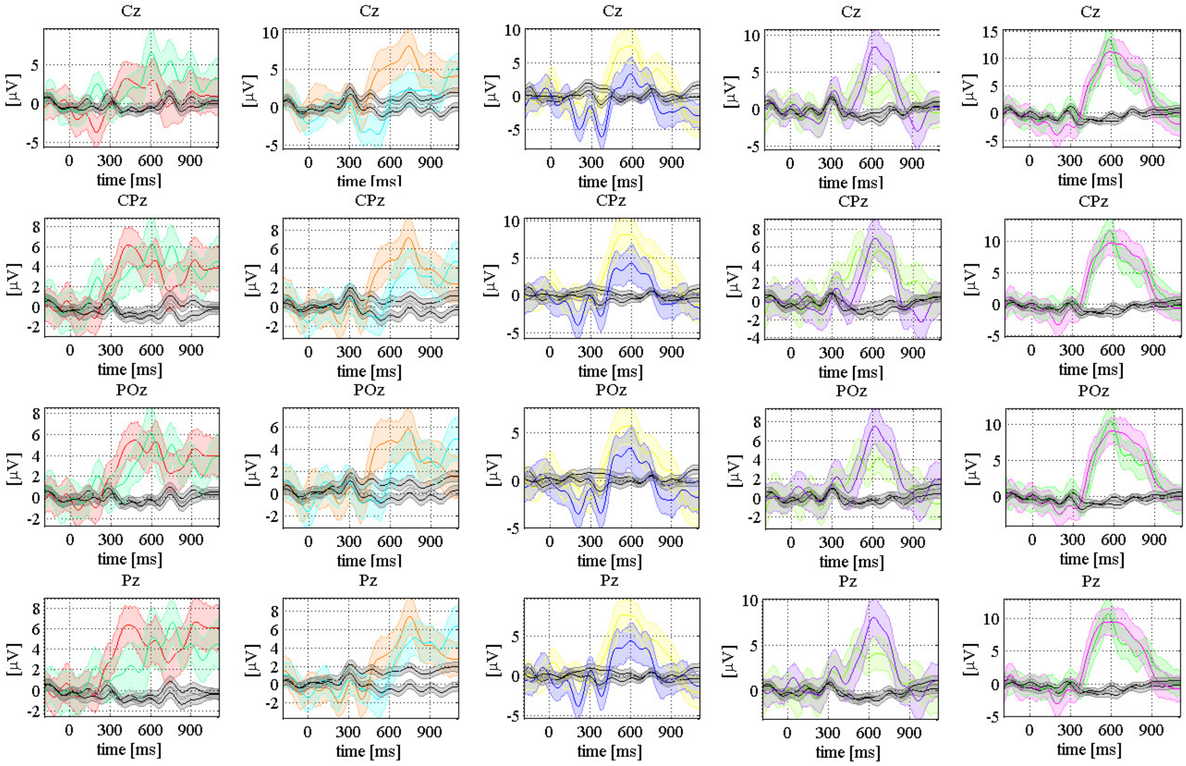

Fig. 2. Grand mean averaged EEG ERP responses together with standard error bars for all the participating users in the study plotted for representative electrodes separately for each direction utilizing horizontal placement. Each column shows $-90°$, $-45°$, $0°$, $45°$, and $90°$ from left to right, in case $0°$ was set to median plane.

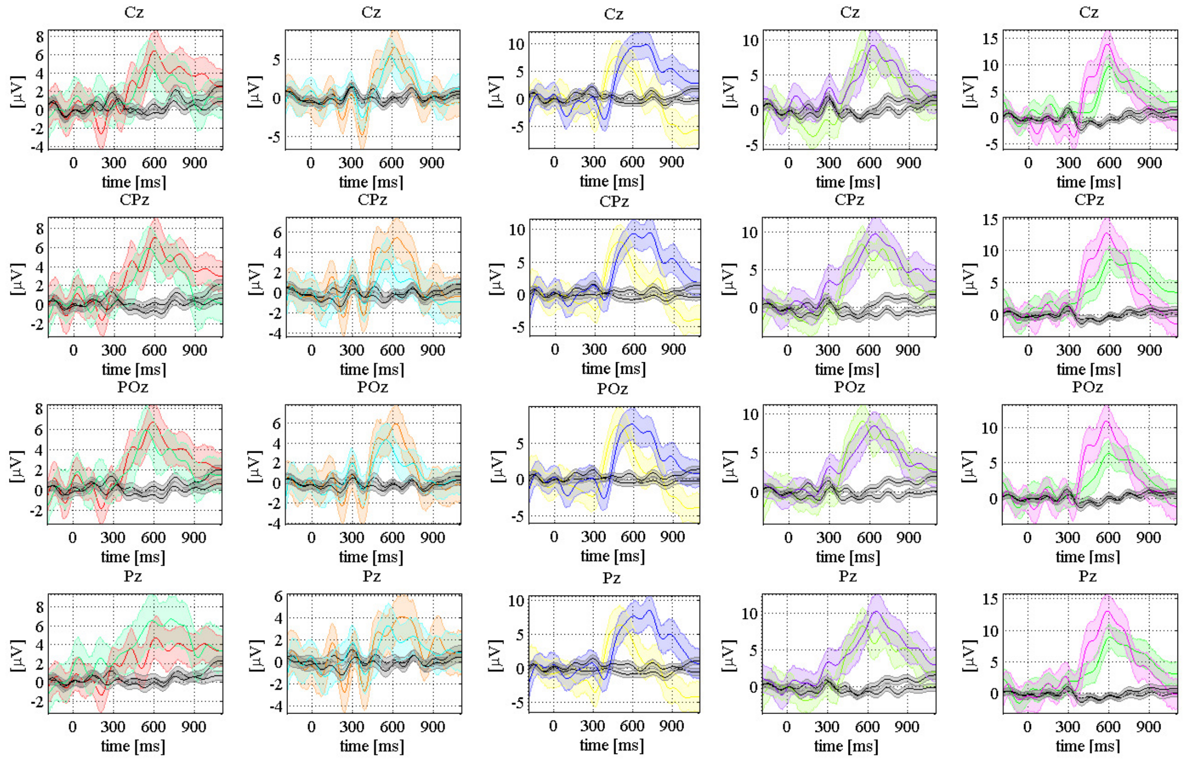

Fig. 3. Grand mean averaged EEG ERP responses together with standard error bars for all the participating users in the study plotted for representative electrodes separately for each direction utilizing elevation used modality. Each column shows $-90°$, $-45°$, $0°$, $45°$, and $90°$ from left to right, in case $0°$ was set to median plane.

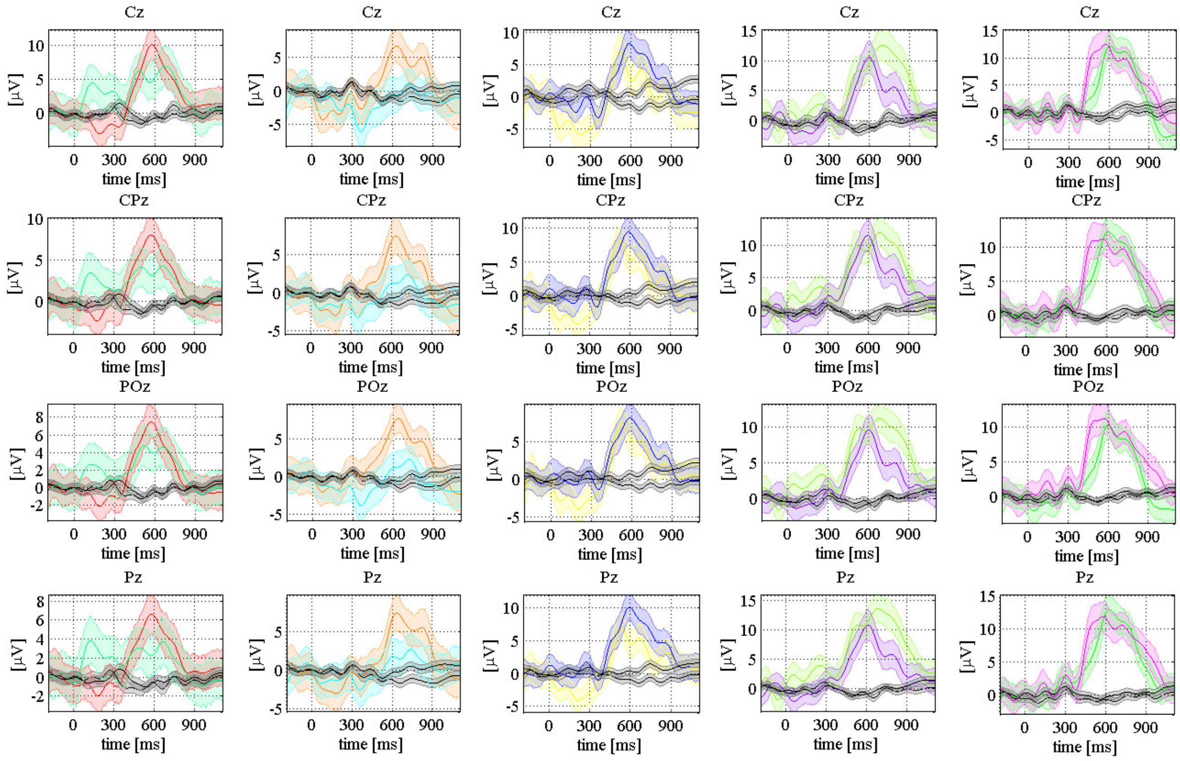

Fig. 4. Grand mean averaged EEG ERP responses together with standard error bars for all the participating users in the study plotted for representative electrodes separately for each direction utilizing elevation and frequency modality. Each column shows $-90°$, $-45°$, $0°$, $45°$, and $90°$ from left to right, in case $0°$ was set to median plane.

features drawn from the $0 \sim 800$ ms ERP interval, with removal of the least significant input features, having $p > 0.15$, and with the final discriminant function restricted to contain a maximum of 60 features.

III. Results

The results of the saBCI EEG experiment have been depicted in Figures 2, 3, and 4 for each modality respectively. Each column in the above figures represents grand mean averaged ERP results at representative four electrodes for each azimuth of presented sound. Each color line depicts the grand mean averaged ERPs with standard error intervals to the *target* command as in Figure 5 respectively, while black lines are the *non–targets*. Eye blink artifacts were removed with 80 μV thresholding. The clear P300 responses were occuring in latency ranges of $400 \sim 1000$ ms in each modality.

As shown in Figure 3, there was less variation between high and low elevation commands in the elevation use modality. Comparing the Figures 3 and 4, the high elevation commands could evoke stronger P300 responses at $-45°$ and $45°$ in the elevation and frequency use modality.

Figure 6 shows a histogram of the classification accuracy of all users with standard deviation (STD) values. Elevation use modality resulted with the lowest dispersion among the three tested modalities.

We also analyzed the classification accuracies of all commands. Summarized results for all subjects have been depicted in Figures 7, 8, and 9. The confusion rates were calculated as the classification results of the P300 responses obtained with the SWLDA classifier application in BCI2000 experimental environment.

Although there was no tendency to mistake only in a certain command utilizing horizontal placement, there were difference of the high accuracy command and low accuracy command for each modality. In case of elevation use modality, the overall accuracy was the best among the three tested modalities. However command "*sa*" and "*ma*" scored lower than the other

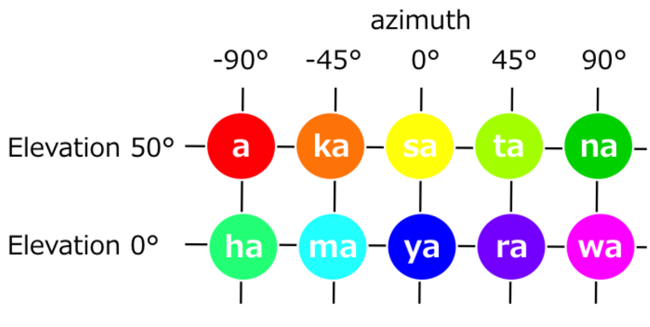

Fig. 5. Target commands placement for each azimuth and elevation. Each color shows the different command with the spatial location.

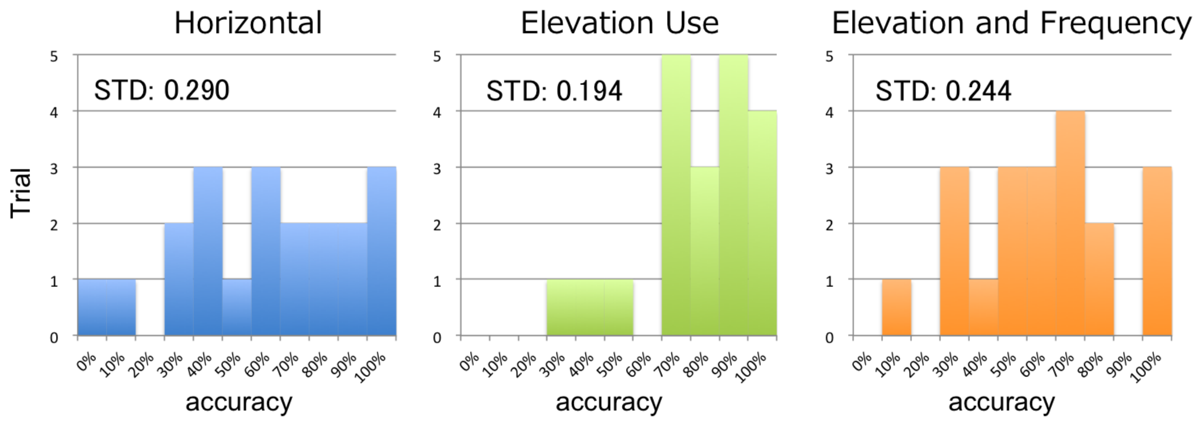

Fig. 6. Histograms of the classification accuracies of all subjects shown for each modality separately.

evaluated commands.

IV. Conclusions

The presented EEG results confirmed the P300 responses feasibility among the experienced and naive saBCI users. The proposed spatial sound stimulus settings for various elevations obtained using HRIR were effective for saBCI paradigm. Elevation use modality resulted with the lowest BCI accuracy dispersion among the tested users and the BCI commands' scores.

Nevertheless, current study is not yet ready to compete with the faster visual BCI spellers due to a lower information transfer rate caused by longer sound stimuli presentation latencies. We plan to extend the proposed saBCI to realize the full set Japanese kana characters–based speller utilizing multi–step spalling procedure and to improve it for a comfortable online spelling.

References

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

[2] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "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, no. 1, pp. 90–100, 2009.

[3] D. Gelinas, "Living with ALS - managing your symptoms and treatment," Online brochure, The ALS Association, 2007. [Online]. Available: http://www.alsa.org/assets/pdfs/brochures/alsa_manual3.pdf

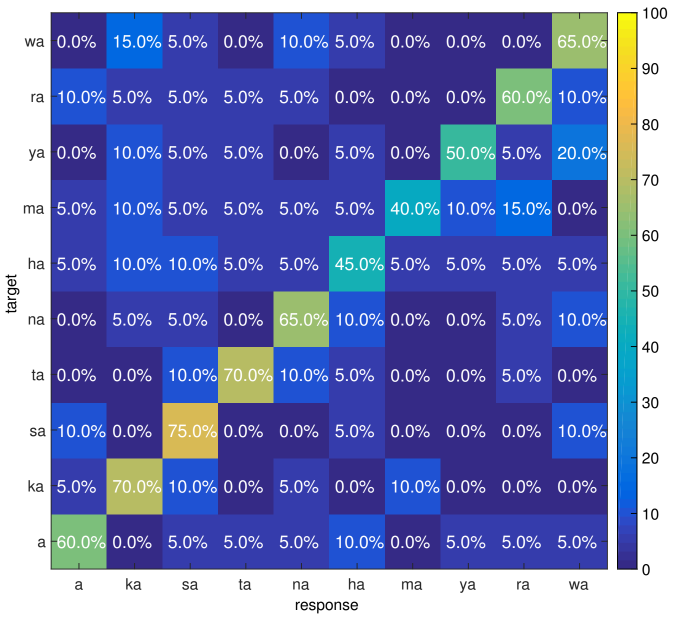

Fig. 7. Classification accuracies in form a confusion matrix from the online EEG experiments depicting averaged results of all users in horizontal modality case. The numbers indicate the averaged accuracies, which have been additionally depicted with color coding. The vertical axis indicates instructed *targets* while the horizontal one the user responses decoded from EEG.

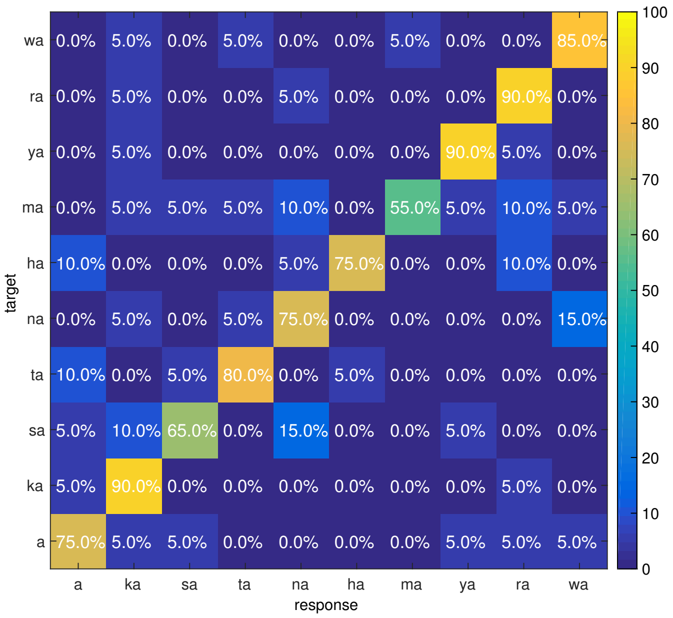

Fig. 8. Classification accuracies in form a confusion matrix from the online EEG experiments depicting averaged results of all users in elevation use modality case. The numbers indicate the averaged accuracies, which have been additionally depicted with color coding. The vertical axis indicates instructed *targets* while the horizontal one the user responses decoded from EEG.

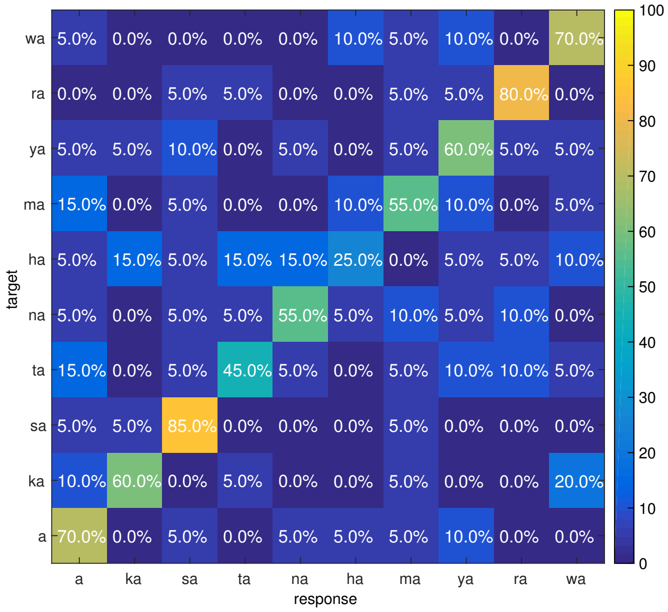

Fig. 9. Classification accuracies in form a confusion matrix from the online EEG experiments depicting averaged results of all users in elevation and frequency modulation modality case. The numbers indicate the averaged accuracies, which have been additionally depicted with color coding. The vertical axis indicates instructed *targets* while the horizontal one the user responses decoded from EEG.

[4] J. R. Patterson and M. Grabois, "Locked-in syndrome: a review of 139 cases." *Stroke*, vol. 17, no. 4, pp. 758–64, 1986. [Online]. Available: http://stroke.ahajournals.org/content/17/4/758.abstract

[5] 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.

[6] 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.

[7] 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

[8] H. Higashi, T. M. Rutkowski, Y. Washizawa, A. Cichocki, and T. Tanaka, "EEG auditory steady state responses classification for the novel BCI," in *Proceedings of the 33rd Annual International Conference of the IEEE EMBS*. Boston MA, USA: IEEE Press, August 30 - September 3, 2011, pp. 4576–4579.

[9] N. Nishikawa, Y. Matsumoto, S. Makino, and T. M. Rutkowski, "The spatial real and virtual sound stimuli optimization for the auditory bci," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. IEEE, 2012, pp. 1–9.

[10] Y. Matsumoto, S. Makino, K. Mori, and T. M. Rutkowski, "Classifying P300 responses to vowel stimuli for auditory brain-computer interface," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, 2013, pp. 1–5, paper ID 388. [Online]. Available: http://www.apsipa.org/proceedings_2013/papers/388_Classifying-P300-Matsumoto-2933943.pdf

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

[12] C. Nakaizumi, K. Mori, T. Matsui, S. Makino, and T. M. Rutkowski, "Auditory brain-computer interface paradigm with head related impulse response-based spatial cues," in *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on*. IEEE, 2013, pp. 806–811.

[13] S. Amano, S. Sakamoto, T. Kondo, and Y. Suzuki, "Development of familiarity-controlled word lists 2003 (fw03) to assess spoken-word intelligibility in japanese," *Speech Communication*, vol. 51, no. 1, pp. 76–82, 2009.

[14] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[15] 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.

[16] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 3933–3936.

[17] 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

[18] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller," *Journal of Neural Engineering*, vol. 3, no. 4, p. 299, 2006. [Online]. Available: http://stacks.iop.org/1741-2552/3/i=4/a=007