

# Spatial Auditory BCI/BMI Paradigm - Multichannel EMD Approach to Brain Responses Estimation

T.M. Rutkowski\*, T. Tanaka†\*, Q. Zhao\* and A. Cichocki\*

\* Laboratory for Advanced Brain Signal Processing & RIKEN Brain-TOYOTA Collaboration Center  
Wako-shi 351-0198 Saitama Japan

E-mail: tomek@brain.riken.jp.jp Tel: +81-48-462-1111 ext. 6993

† Tokyo University of Agriculture and Technology, Tokyo 184-8588 Japan  
E-mail: tanakat@cc.tuat.ac.jp Tel/Fax: +81-42-388-7439

**Abstract**—A novel spatial auditory BCI/BMI paradigm is proposed which is based on responses to 7.1 channels surround sound audio stimuli. The approach is based on a monitoring of brain electrical activity by means of the electroencephalogram (EEG) with utilization of a multichannel EMD technique. Owing to its non-invasive nature, the EEG based BCI/BMI are envisaged to be at the core of future intelligent prosthetic. A spatial auditory stimulus is a very interesting and not mentally demanding paradigm receiving recently more attention in computational neuroscience applications due to involvement of a secondary in human-machine-interaction auditory modality. We propose to utilize spatial audio stimuli design and application in new BCI/BMI paradigms where users intentionally direct their attention to different locations in surround sound environment with steady-state tonal frequency stimuli.

## I. INTRODUCTION

Brain computer and brain machine interfaces (BCI/BMI) [1] or generally responses to spatial audio stimuli are typically based on the monitoring of brain electrical activity by means of the electroencephalogram (EEG) [2]. Owing to its non-invasive nature, the EEG based BCI/BMI are the best candidates to be at the core of future “intelligent” prosthetic devices. They are particularly suited to the needs of the handicapped as well the core of smart environments, and for users awareness or attentional levels estimation [3]. The development of such devices, however, is very demanding since the brain responses to the same stimulus not only depend on the current mental, especially emotional, state of the user but also it varies across the users as well they are very often degraded by much stronger in power muscle activity electrical noise interference.

A concept of a spatial auditory stimulus creates a very interesting possibility to target “a less crucial” auditory activity, which is not as critical as vision during operation of machinery or driving a car. Auditory BCI/BMI is thus potentially a less mentally demanding paradigm receiving recently more attention in computational neuroscience applications [2], [4]. We propose to utilize spatial audio stimuli design with a target application in a new BCI/BMI paradigms where users intentionally direct their attention to different locations in surround sound environment with various tonal frequency stimuli [5] as depicted in Figure 1. Contemporary applications limit their scope to frontal surround sound speakers [6], while our proposal includes also rear speakers sound presentation

allowing for seven commands BCI/BMI applications (eight in case of octagonal speakers setups).

In order to identify user’s target responses to presented spatial stimuli we first have to preprocess the EEG signals in order to decompose them into components carrying stimuli evoked potentials (so called event-related potentials - ERP). In order to achieve it we propose a signal processing pipeline composed of EMD technique with spectral clustering followed by signals averaging and their locations estimation as discussed in following sections.

## II. METHODS

In the proposed approach, we analyze responses from experiments based on a spatial tonal stimuli which were conducted

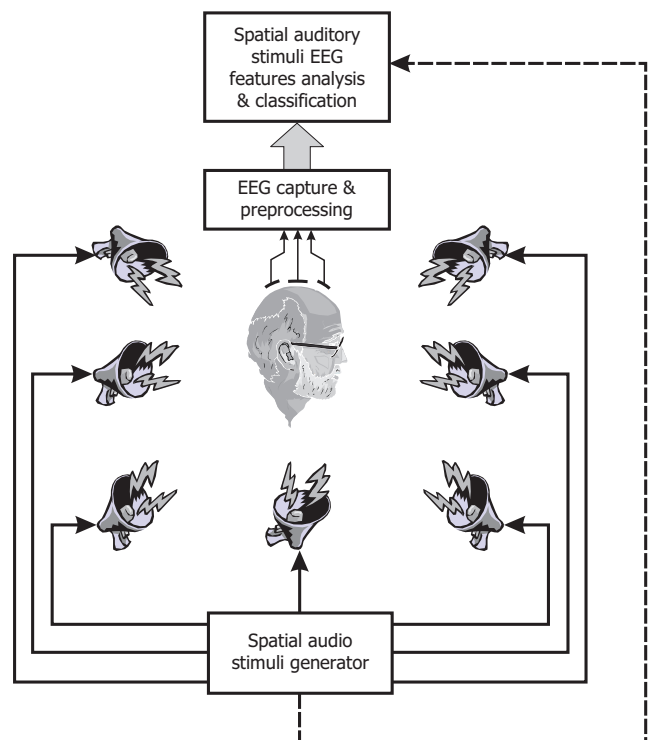


Fig. 1. Spatial audio experimental setup. The subject’s EEG is synchronously recorded during the experiment.

in the Laboratory for Advanced Brain Signal Processing, BSI RIKEN, within a surround sound 7.1 channels system as presented in Figure 1, where four subjects were positioned in the middle of the speakers systems and requested to direct attention to single direction speakers and ignoring the others in each trial. The target and non-target direction sequences were presented randomly in 1/7 ratio.

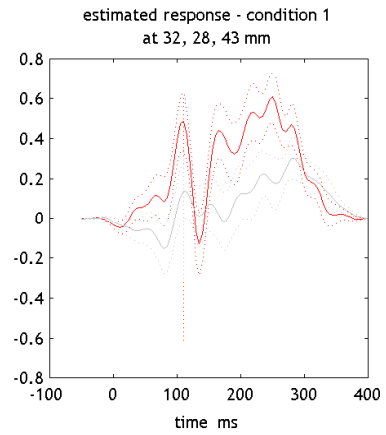
Within this framework, the subjects are asked to focus their attention on a direction of the tonal or environmental sound. The EEG responses are recorded with a BIOSEMI system with sampling frequency of 2048Hz and 128 electrodes placed on a head as in extended 10/20 EEG system [7]. Additionally vertical and horizontal eye-movements were recorded in order to have a reference signal indicating potential muscle activity reference used later in an EMD algorithm.

We utilize previously proposed by the authors an approach utilizing empirical mode decomposition (EMD) technique combined with frequency domain clustering scheme as in [8]. This way, the multichannel and multimodal signal decomposition technique uses the EEG captured by several electrodes located over auditory and temporal brain areas, subsequently preprocessed, and transformed into informative time-frequency traces, which very accurately visualize frequency and amplitude with utilization of an extension of the EMD [9], [8]. This approach allows us to separate from all EEG channels only the components carrying brain activities with frequencies as in ASSR stimuli signals to which subjects attend in spatial environment. The so obtained brain activity spatial patterns clearly follow the expectations of stronger activities in auditory and temporal cortical areas related to the attended sound directions.

EMD utilizes empirical knowledge of oscillations intrinsic to a signal in order to represent them in a form of a superposition of components, called *intrinsic mode functions* (IMF). IMFs are characterized by well defined instantaneous frequencies. To obtain an IMF from a single channel EEG it is necessary to remove the local riding waves and asymmetries. They are estimated from local envelope of minima and maxima of the waveform. The technique of finding IMFs corresponds to eliminating riding-waves from the signal. It ensures that the instantaneous frequency will have no fluctuations caused by an asymmetric waveform. In each cycle, the IMF is defined by zero crossings and involves only one mode of oscillation, thus not allowing complex riding waves. Notice that an IMF is not limited to be a narrow band signal, as it would be in classical Fourier or wavelets decompositions. In fact, an IMF can be both amplitude and frequency modulated simultaneously, as well as non-stationary or non-linear.

EMD decomposes a signal in hand into IMFs [9] and represents in form of “oscillatory modes” which obey the following two conditions:

- (i) the number of extrema and the number of zero crossings should be either equal or differ at most by one;
- (ii) at any point, the envelope defined by the local maxima and the envelope defined by the local minima is zero mean.



PPM at 110 ms (70 percent confidence)  
512 dipoles  
Percent variance explained 94.40 (93.74)  
log-evidence = 2087.4

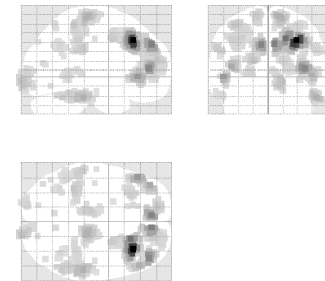
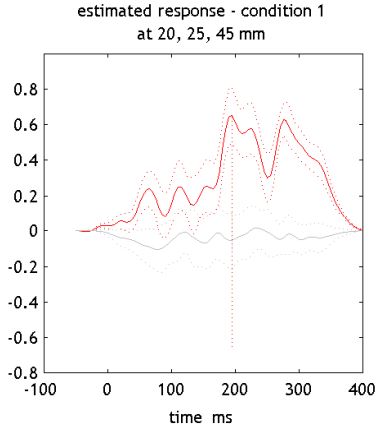


Fig. 2. The results of EEG activity to AM modulated tones (the upper panel of the above figure) mapping at 110ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user’s attention (red - target or expected stimuli direction; gray - non-target or ignored stimuli direction). These results show that already 110ms after stimuli onset subject’s brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources of the visualized ERP as visualized with SPM8 [10].

Since IMF represents an oscillatory mode within a signal its periods, which are defined by zero crossings, correspond to the only *one* mode of oscillation. Both the amplitude and frequency of this oscillation may vary over time, in other words, the oscillation is not necessarily stationary nor narrow-band.

The process of extracting an IMF from a signal  $x(t)$  is defines as “the sifting process” [9] and constitutes of the following steps:

- 1) first determine the local maxima and minima of  $x(t)$ ;
- 2) next generate the upper and lower signal envelope by connecting those local maxima and minima respectively by an interpolation method (e.g., linear, spline, piecewise spline [9], [2]);
- 3) after that determine the local mean  $m_1(t)$ , by averaging the upper and lower signal envelope;
- 4) finally subtract the local mean from the data:  $h_1(t) = x(t) - m_1(t)$ .



PPM at 195 ms (66 percent confidence)  
512 dipoles  
Percent variance explained 94.40 (93.74)  
log-evidence = 2087.4

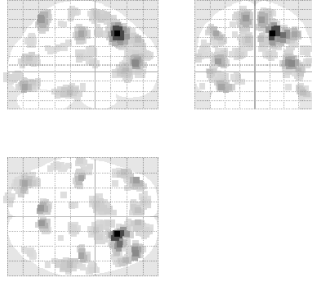
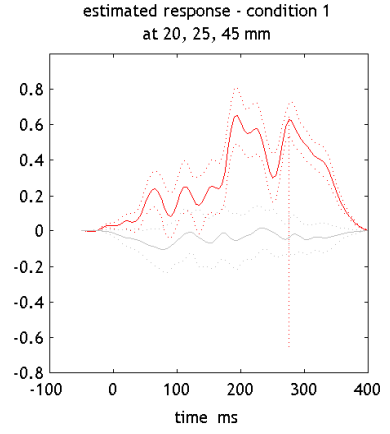


Fig. 3. The results of EEG activity to AM modulated tones (the upper panel of the above figure) mapping at 200ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user’s attention (red - target or expected stimuli direction; gray - non-target or ignored stimuli direction). These results show that already 200ms after stimuli onset subject’s brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources of the visualized ERP as visualized with SPM8 [10].

Ideally,  $h_1(t)$  satisfies the conditions (i) and (ii) of an IMF, however, typically this procedure needs to be repeated until the first IMF is extracted. In order to obtain the second IMF one applies the sifting process to the residue  $\epsilon_1(t) = x(t) - \text{IMF}_1(t)$ , obtained by subtracting the first IMF from  $x(t)$ ; the third IMF is in turn extracted from the residue  $\epsilon_2(t)$  and so on. The decomposition is completed once two consecutive sifting results are similar; the empirical mode decomposition of the signal  $x(t)$  may be thus written as:

$$x(t) = \sum_{k=1}^n \text{IMF}_k(t) + \epsilon_n(t), \quad (1)$$

where  $n$  is the number of extracted IMFs, and the final residue  $\epsilon_n(t)$  is either the mean trend or a constant. Note that the IMFs are not guaranteed to be mutually orthogonal, but are often close to orthogonal; note also noteworthy that IMFs are adaptive, that is, two independent realizations of a signal with the same statistics may have a different number of IMFs.



PPM at 275 ms (90 percent confidence)  
512 dipoles  
Percent variance explained 94.40 (93.74)  
log-evidence = 2087.4

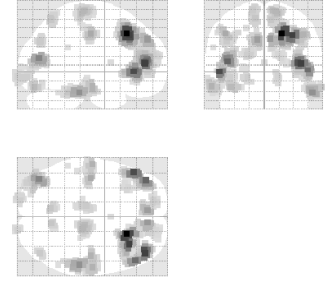
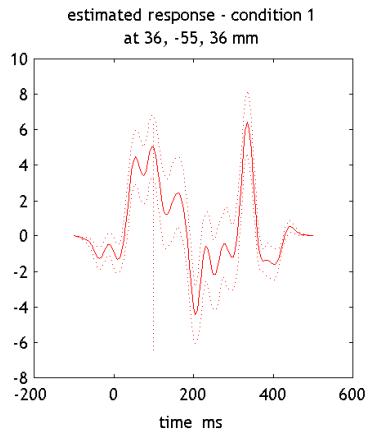


Fig. 4. The results of EEG activity to AM modulated tones (the upper panel of the above figure) mapping at 275ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user’s attention (red - difference between target and non-target). These results show that already 200ms after stimuli onset subject’s brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources differential activity of the ERP as visualized with SPM8 [10].

In order to compare all IMFs extracted from the analyzed channels (two EOG and 128 EEG in this paper) we propose to transform them to Hilbert domain in order to capture spectral content carried by all of them. The spectral amplitude ridges (traces of amplitude in Hilbert domain) are further treated as features and compared for their similarity.

We use correlations between variables as “a distance measure” in order to capture spectral similarity across the IMFs. Once cross-correlation analysis is performed for all IMFs from all analyzed channels a hierarchical cluster analysis (HCA) using a set of dissimilarities for the  $n$  objects to be clustered is applied. In the HCA initially, each vector representing amplitude ridges values is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters. Such procedure continues until there is just a single cluster. At each stage distances between clusters are recomputed by the Lance–Williams dissimilarity update formula with a single linkage method clustering method. The single linkage method is



PPM at 100 ms (76 percent confidence)  
512 dipoles  
Percent variance explained 95.45 (94.99)  
log-evidence = -2358.5

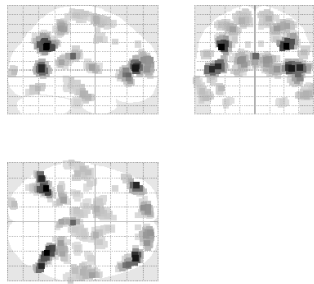
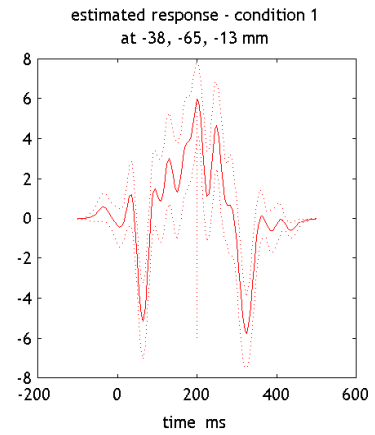


Fig. 5. The results of EEG activity to environmental sound (the upper panel of the above figure) mapping at 100ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user's attention (red - difference between target and non-target). These results show that already 110ms after stimuli onset subject's brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources differential activity of the ERP as visualized with SPM8 [10].

closely related to the minimal spanning tree concept and it adopts a "friends of friends" strategy for clustering [11]. The first set is for distances below median of all distances and those components from different channels are classified as similar and originating from very strong muscle activity interference. The remaining set of clusters with distances above the median represent IMFs carrying neurophysiological signals only. Note that the proposed method saves brain-activity-related low frequency spectral content of neurophysiological signals under very strong EOG interference.

Finally the reconstructed neurophysiological signals are averaged to compare target and non-target evoked potentials as visualized in Figures 2, 3 and 4. The results presented there show that it is possible to discriminate brain evoked responses with highest differences for 110ms, 200ms and 275ms discriminating responses to target (expected) spatial sound locations (red lines) and the ignored ones (gray lines). We also present simple grand mean averages for four subjects for target and non-target directional stimuli as presented in



PPM at 200 ms (72 percent confidence)  
512 dipoles  
Percent variance explained 95.45 (94.99)  
log-evidence = -2358.5

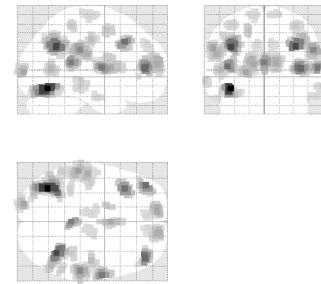


Fig. 6. The results of EEG activity to environmental sound (the upper panel of the above figure) mapping at 200ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user's attention (red - difference between target and non-target). These results show that already 200ms after stimuli onset subject's brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources differential activity of the ERP as visualized with SPM8 [10].

Figures 8 through 12.

### III. DISCUSSION

The obtained results have shown that it is possible to preprocess noisy EEG signals (mostly contaminated by eye movements, blinks, etc.) utilizing the proposed multichannel EMD extension with spectral clustering to find components carrying stronger activities related to attended sound direction. The comparison of results presented in Figures 2 through 7 for a single subject listening to spatial tonal and environmental sound stimuli guided the analysis further to choose little more difficult (according to subjects reports) tonal stimuli for which responses were more stable within EEG. ERP to tonal stimuli were used to further calculate grand mean averages of four subjects in order to present a stable trend across different brains.

The grand mean average of four subjects presented in Figures 8, 9 and 10 with results from the vertex, occipital

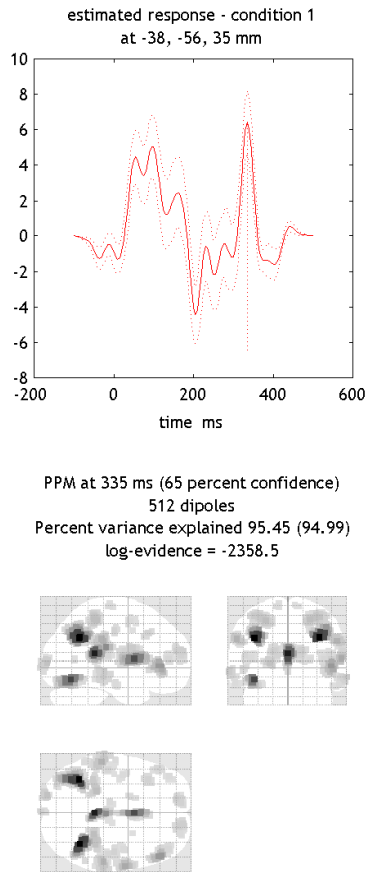


Fig. 7. The results of EEG activity to environmental sound (the upper panel of the above figure) mapping at 335ms after stimulus onset in response to spatial audio stimuli showing possibility to discriminate spatial activity maps based on the user's attention (red - target or expected stimuli direction; gray - non-target or ignored stimuli direction). This results shows that already 200ms after stimuli onset subject's brain reacts differently and this could be already captured by EEG. The lower panel of the figure presents the estimated sources of the visualized ERP as visualized with SPM8 [10].

and parietal cortical regions confirmed existence of a P300 response [7] in spatial audio stimuli oddball paradigm.

The more interesting results from frontal cortical areas are presented also as grand mean averages of the same four subjects in Figures 11 and 12 where differences in ERP responses were found in early time ranges of 150 – 300ms.

#### IV. CONCLUSIONS

We presented a proposal to create the spatial sound directed attention paradigm as a candidate for BCI/BMI technologies which in comparison to contemporary application utilizes also the rear direction (behind the head) speakers to introduce more possible commands.

This is a step forward in EEG signal processing applications which could be useful primarily for creating novel and user friendly brain-machine-interfaces that would be flexible, adaptive and response automatic based on the detection of subject's spatial auditory focused attention, thus resulting in

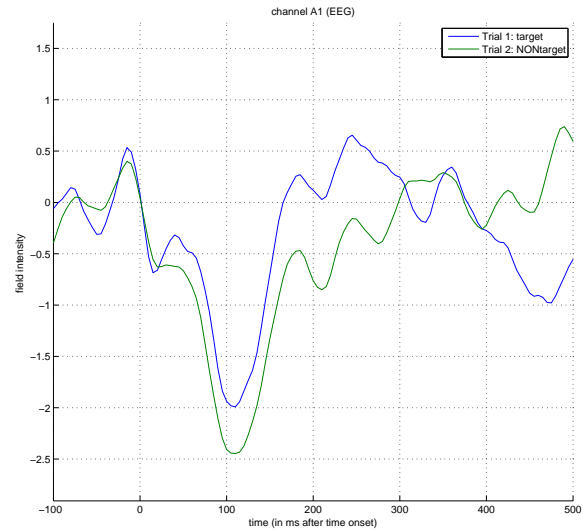


Fig. 8. Grand mean average of ERPs result for four subjects listening to spatial AM modulated tones (stimuli presented over 200ms with 500ms repetition rate) targets (blue) and non-targets (green) for the vertex electrode A1 ( $C_z$  in 10/20 EEG systems) showing clear difference in responses in the range of 200 – 300ms.

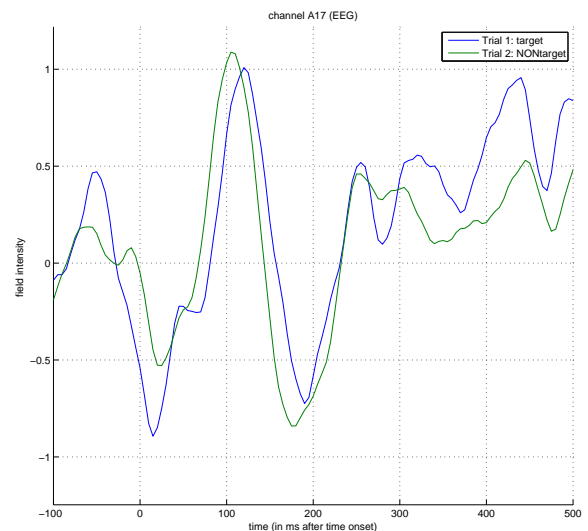


Fig. 9. Grand mean average of ERPs result for four subjects listening to spatial AM modulated tones (stimuli presented over 200ms with 500ms repetition rate) targets (blue) and non-targets (green) for the parietal electrode A17 (close to  $P_1$  in 10/20 EEG systems) showing clear difference in responses in the range of 250 – 450ms.

fast estimation of user's intention in relation to the presented spatial stimuli.

#### ACKNOWLEDGMENTS

Authors would like to thank Sungyoung Kim and Masahiro Ikeda of YAMAHA Corporation for their technical support and rental of an experimental spatial auditory equipment used for the experiments.

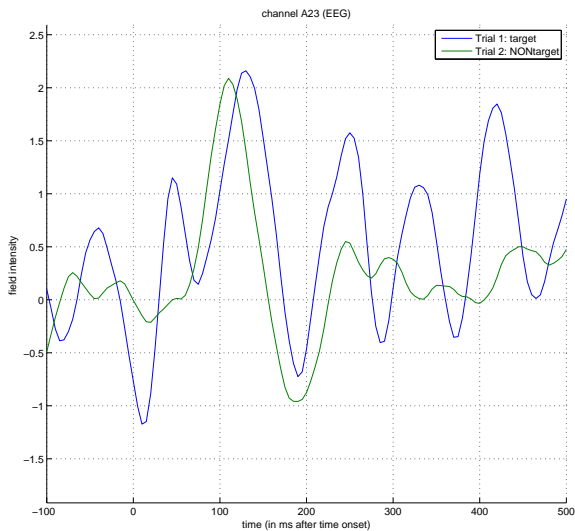


Fig. 10. Grand mean average of ERPs result for four subjects listening to spatial AM modulated tones (stimuli presented over 200ms with 500ms repetition rate) targets (blue) and non-targets (green) for the occipital electrode A23 (close to  $Oz$  in 10/20 EEG systems) also showing clear difference is responses in the range of 250 – 450ms similarly as in Figure 9.

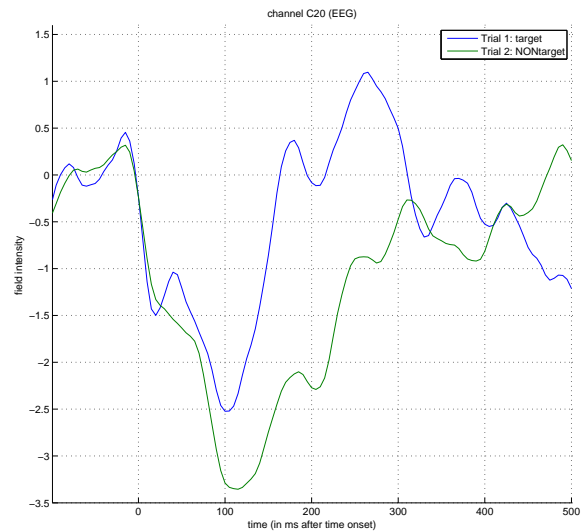


Fig. 12. Grand mean average of ERPs result for four subjects listening to spatial AM modulated tones (stimuli presented over 200ms with 500ms repetition rate) targets (blue) and non-targets (green) for frontal electrode C20 ( $Fz$  in 10/20 EEG systems) also showing clear difference is responses also in the very early range of 150 – 300ms (compare Figure 11).

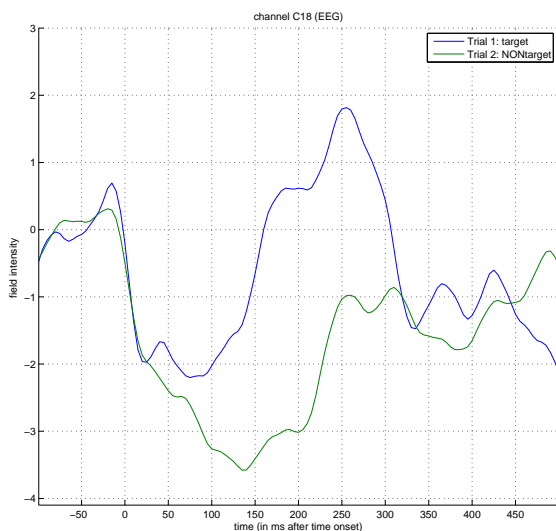


Fig. 11. Grand mean average of ERPs result for four subjects listening to spatial AM modulated tones (stimuli presented over 200ms with 500ms repetition rate) targets (blue) and non-targets (green) for forehead electrode C18 ( $FPz$  in 10/20 EEG systems) also showing clear difference is responses in the very early range of 150 – 300ms. So early in time and clear response difference suggest the forehead as a perfect BCI/BMI location for the presented paradigm.

This research was supported in part by KAKENHI, the Japan Society for the Promotion of Science grant no. 21360179.

#### REFERENCES

[1] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A.-H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, “Noninvasive BCIs:

Multway signal-processing array decompositions,” *Computer*, vol. 41, no. 10, pp. 34–42, 2008.

[2] T. M. Rutkowski, A. Cichocki, and D. Mandic, “Information fusion for perceptual feedback: A brain activity sonification approach,” in *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (D. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, eds.), pp. 261–273, Springer US, 2008.

[3] T. M. Rutkowski, A. Cichocki, A. L. Ralescu, and D. P. Mandic, “Emotional states estimation from multichannel EEG maps,” in *Advances in Cognitive Neurodynamics ICCN 2007 Proceedings of the International Conference on Cognitive Neurodynamics* (R. Wang, F. Gu, and E. Shen, eds.), Neuroscience, pp. 695–698, Springer Berlin & Heidelberg, 2008.

[4] T. Rutkowski, D. Mandic, and A. Barros, “A multimodal approach to communicative interactivity classification,” *The Journal of VLSI Signal Processing*, vol. 49, no. 2, pp. 317–328, 2007.

[5] 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), p. P5, November 11–13, 2009.

[6] 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, p. e9813, 04 2010.

[7] E. Niedermeyer and F. L. Da Silva, eds., *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 5 ed., 2004.

[8] T. M. Rutkowski, A. Cichocki, T. Tanaka, D. P. Mandic, J. Cao, and A. L. Ralescu, “Multichannel spectral pattern separation - an EEG processing application,” in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2009)*, pp. 373–376, IEEE, 2009.

[9] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, March 1998.

[10] Wellcome Trust Centre for Neuroimaging, “Statistical parametric mapping - SPM8 package.” <http://www.fil.ion.ucl.ac.uk/spm/>, 2010.

[11] F. Murtagh, “Multidimensional clustering algorithms,” *COMPSTAT Lectures*, vol. 4, 1985.