DETECTING ATTENTION SHIFT FROM NEURAL RESPONSE BASED ON BEAT-FREQUENCY-MODULATED MUSICAL EXCERPTS

Takashi G. Sato, Yoshifumi Shiraki, Takehiro Moriya

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
Atsugi, Kanagawa, Japan
takashi_goto_sato@ieee.org

ABSTRACT

This paper presents a new approach for detecting attention from auditory steady-state responses (ASSR) by using musical excerpts. The feature extraction process for electroencephalogram (EEG) signal is combined with a support vector machine as a binary discriminator. A novel modulation that emphasizes the beat timing of the excerpts has enhances the EEG response. Thanks to the beat-locked epoch extraction and additional information that signals the locked excerpt, the estimation errors are less than 8% using only ten seconds of data. Contrary to our expectations, a waveform-averaging method outperforms a harmonic filter bank and a bin-subtraction methods in the frequency domain for feature extraction. Overall, attention estimation from EEGs using musical excerpts as stimuli has been successfully achieved, which represents significant progress towards the development of a mass-EEG measurement system.

Index Terms— Brain-computer interface, musical excerpt, beat, auditory steady state response, support vector machine.

1. INTRODUCTION

The brain-computer interface (BCI) is a developing technology aimed at controlling an external device by measuring the user's brain activity. The major target of BCIs has been disabled persons, such as those with amyotrophic lateral sclerosis. Thanks to recent developments in electroencephalography technology, BCIs are no longer expensive nor restricted to laboratory use [1]. Moreover, it is becoming possible to gather physiological data from large audiences attending exhibitions or live performances [2]. This situation opens up new possibilities for BCI technology. For example, if we can determine the object on which an audience is focusing (like a specific exhibition booth or specific sound stream), we can use such information to evaluate the object's appearance or as feedback to enhance audience experiences (Fig. 1). Our final goal is to achieve such mass-electroencephalogram (EEG) measurement. In this sense, though we use the term BCI, our aim is more at detecting the object that people are focusing on than on maximizing the transfer rate of their intention. For this purpose, a simple EEG device is preferable to one that uses dozens of channels.

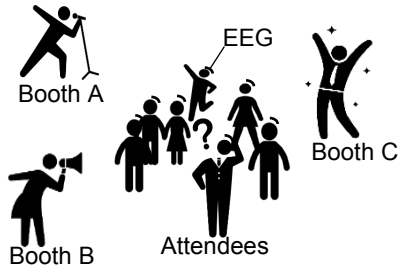


Fig. 1. Concept of mass-EEG measurement. In this example, each booth has its own modulation stimulus. We ascertain the booth on which attendees are focusing from their EEGs. The results are used to evaluate the booths and as feedback to enhance the attendees' experience.

Although we can use a visual or auditory stimulus for cueing, we choose an auditory one as our first step. This is because visual stimuli that evoke visually steady-state responses (VSSRs) are relatively more annoying than auditory stimuli. Besides, audio steady-state responses (ASSRs), which were once considered to have little hope for detecting attention, are now showing some optimistic results for attentional detection [3-5]. Traditionally, a pure tone modulated with frequencies from 10 to 40 Hz or event-related potential (ERP) has been used [6-8]. Recently, however, repetitive stimuli whose frequency ranges from 0.5 to around 5 Hz are attracting more interest because they make a distinct difference from the pure ERP [9]. Several studies have suggested that there are beat-locked responses in EEGs and that the responses may change according to the attentional level of the participants to the stimulus [10, 11]. Moreover, a study showed that presented excerpts can be estimated from EEGs, though it requires offline analysis using with the whole set of EEG data [12].

However, BCIs using this frequency range (0.5 to 5Hz) have rarely been studied. Although musical excerpts are expected to ease the difficulty of attentional tasks, most studies have used pure tones or artificial sounds. In this study, we focused on this low-frequency range and used musical excerpts as stimuli. Since beat timing is close to the frequency range of interest, we created frequency-modulated stimuli by emphasizing the beat timing in the excerpts. Note that our aim is not to design a BCI speller but rather to determine the object of an audience's attention in a common situation. Our study starts with a dual listening task, and the analysis is more aimed to run with a few electrode channels.

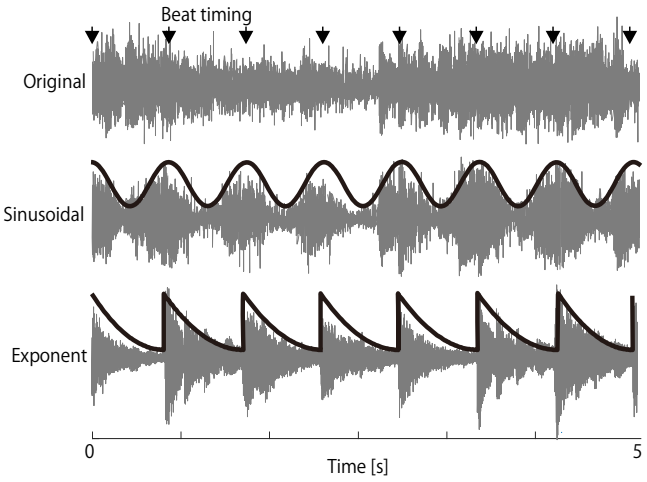


Figure 2. Examples of modulated sound. The thick black line shows the modulation coefficient calculated from (1) and (2).

2. METHODS

2.1. Stimuli

As shown in Table 1, stimuli consisted of artificially made stimuli [4] and musical excerpts. All musical excerpts were sampled at 44.1 kHz and lasted for 50 seconds. They were modulated in amplitude with the frequency matched to that of the beat. Since the original excerpts were recorded ones, there were fluctuations in tempo. Therefore, we adjusted modulations to match the beats as follows. For sinusoidal modulation:

$$\boldsymbol{m_s(t) = A\cos\left(2\pi\left(\frac{t-B_n}{B_{n+1}-B_n}\right)\right)+(1-A).} \quad (1)$$

And for exponential modulation:

$$\boldsymbol{m_e(t) = A\left(1-\left(\frac{t-B_n}{B_{n+1}-B_n}\right)\right)^2+(1-A).} \quad (2)$$

Here, t is time, B_n is the time of the n th beat, and A is the modulation depth parameter. We used A=0.8 in our experiment. Examples of the modulated sounds are shown in Fig. 2. These modulations were conducted to enhance the ASSR by emphasizing the beat rhythm. No modulation was done for A514 and A824 since they are already modulated.

2.2. Participants and tasks

A preliminary test was conducted to see difference in the EEG responses caused by different modulations. In this test, sinusoid modulation and exponential modulation were applied to the excerpts and presented monaurally to the participants.

In the main experiment, a block consisted of monaural trials (only one modulated excerpt presented) and binaural trials (two modulated excerpts presented, one played from the right and the other from the left). In the binaural trials, participants were asked to focus on one of the streams (left or right), which was indicated by the experimenter, and asked to count the number of beats. The task was given to make sure that the participants were focusing on the indicated stream. After the trial ended, they were asked to report the number of beats they had counted and evaluate on a five-point scale the easiness of keeping their focus on the stream. Excerpts presented from the left and right were counterbalanced and presented in random order. Artificial sounds and music excerpts were tested separately. Since conditioning trials (pink noise presented monaurally) were included, the block totals to 22 trials.

Table 1. List of excerpts. Four different type of musical excerpts were selected. The selections had different beats per minute (BPM) and were different types of music. A514 and A824 were artificially made, similar to [4], with 520-Hz carrying frequency with 1.4-Hz modulation and 800-Hz carrier with 2.4-Hz modulation, respectively.

ID	C112	E801	E802	E805	A514	A824
Bpm	95	176	60	115	84	144
Type	Classic Swinging	Pops Happy	New age Beautiful	Techno Bouncy	520-Hz carrier	800-Hz carrier

The experiments were conducted in an acoustically shielded room. The participants sat on a comfortable chair facing loudspeakers. They were not restricted from blinking, but they were requested to fixate a visual marker and refrain from muscle movement, especially tapping or moving with the beat rhythm. In the middle of the block, the participants could rest for few minutes if they so desired, and more than 15 minutes of rest was allotted between blocks. A total of three blocks were conducted.

We recruited five adults (all females) for the preliminary test and ten adults (three males and seven females) for the main experiment. All the participants were healthy and had no hearing difficulties. They provided written informed consent prior to the experiments. All the experiments were approved by the Ethics and Safety Committee of NTT Communication Science Laboratories and adhered to the tenets of the Helsinki Declaration.

Sound pressure at the participant's position was kept below 80 dB (average 60 dB) for the duration of the experiment. EEGs were recorded using a DSI-24 (wearable sensor) sampled at 300 Hz with 20 electrodes. Custom software written in LabVIEW 2012 (National Instruments) was used for stimulus presentation and EEG device control.

2.3. Data processing

All signal processing was conducted offline using MATLAB R2017a (The Math Works) with EEGLab [13]. After band-pass filtering, artifact rejection was conducted. More specifically, an independent component analysis (ICA) was conducted using the data both from the resting trials and pink noise trials. Then, blinking-related components were specified and removed from the other data.

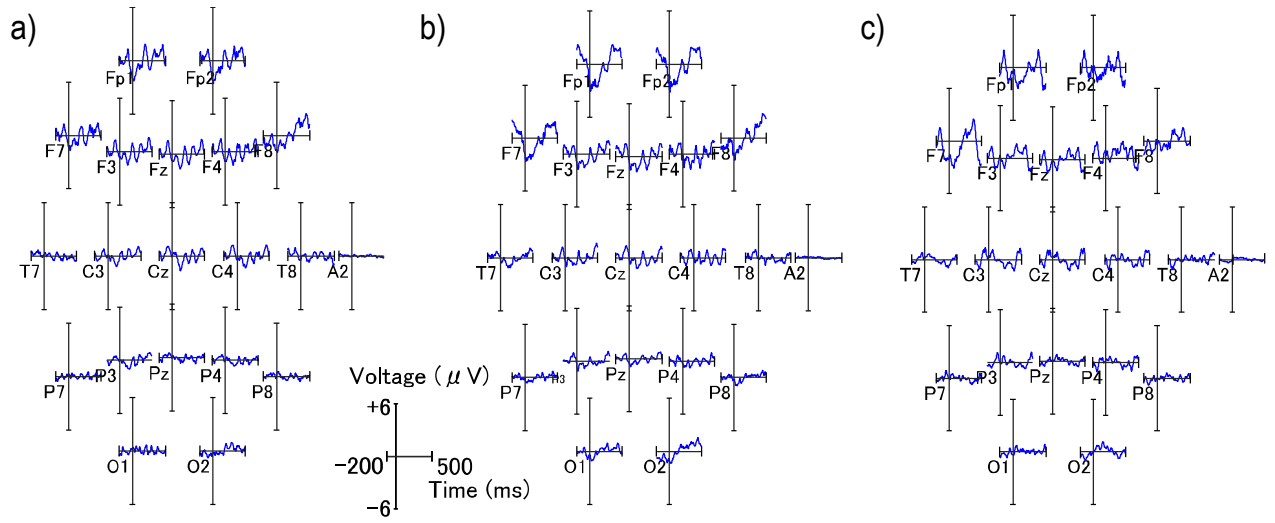


Figure 3. Averaged EEG reponse locked to the event of excerpt E805. a) Without modulation. b) Exponential modulation. c) Sinusoidal modulation.

Our purpose here is to create a reliable binary classifier to estimate the attended auditory object from the EEG data. We can utilize both the frequency and onset of the beat timing as prior knowledge. In our processing, beat timings from different excerpts ($e_{(-1,i)}$, $e_{(1,i)}$) were treated as different events. In other words, after feature vectors were extracted, additional information was given to signal the locked excerpt. A support vector machine (SVM) [14] was used to train and estimate the attended stream. Training was done for each combination of stimulus and participant. Evaluations were made using the ten-fold cross-validation method. For the purpose of seeking good feature representation, we compared three methods.

The first method (waveform) averages the waveform locked to the events. The EEG waveform locked to the *i*th event of excerpt *m* is extracted as epoch

$$\boldsymbol{X}_{(m,i)} = \left\{ \boldsymbol{x}_{e(m,i)+T_s}, \cdots, \boldsymbol{x}_{e(m,i)}, \cdots, \boldsymbol{x}_{e(m,i)+T_e} \right\}, \quad (3)$$

where x_t is EEG output at time t, and T_s and T_e specify the time ranges used for averaging, which were set as -200 ms and 500 ms in this method. High-cut filtering was applied prior to epoch extraction, and downsampling was conducted afterward to preserve the precise timing of the event. The averaged feature vector can be calculated as

$$\mathbf{v} = \left\{ \sum_{i=n}^{n+8} \mathbf{X}_{(m,i)} \right\}. \quad (4)$$

The vector was then normalized, and additional information feature m (represented as -1 or 1) was added to signal the locked excerpt.

The second method (harmonic filter) focuses on the given frequency of the excerpt. In this method, a bank of band-pass filters whose center frequency matches that of the given excerpts was formed and applied to the EEG data. The band-pass filters, f_w, were set to pass plus/minus 1 percent of center frequency w. Since our interest lies in multiples of the basic modulation frequency, the filter bank was applied as

$$\mathbf{F}_{(t)} = \left\{ \boldsymbol{f}_{m_1}(\boldsymbol{x}_t),\ \boldsymbol{f}_{2m_1}(\boldsymbol{x}_t),\ \boldsymbol{f}_{m_2}(\boldsymbol{x}_t),\ \boldsymbol{f}_{2m_2}(\boldsymbol{x}_t) \right\}^{\mathbf{T}}, \quad (5)$$

where m_1 and m_2 are the same as the beat frequency of the presented excerpts. The outputs with 100- and 300-ms latency from each excerpt's events were extracted. After this epoch extraction, the vector was averaged eight times, and, as in the first method, information about the locked excerpt was added.

The third method (bin subtraction) is based on the assumption that the signal amplitude at a given narrow frequency bin should be similar to the signal amplitude of the mean of the surrounding frequency bins if no stimulus exists [9]. According to this assumption, a 0.5- to 30-Hz filtered waveform was epoch extracted as in (3) with [T_s, T_e]=[-4s, 5s]. A Kaiser window (beta=2) was applied, and then 2^{14}-point fast Fourier transform (FFT) was conducted. Target bins (frequencies match the bpm and its multiples) were taken and subtracted using the surrounding bins. This method is a replication of the one in [9, 11] but different in the length of the waveform used for FFT. As far as we know, this is the first time this method has been applied for BCI purposes. In this method, information about locked excerpt was also added as a feature.

Since increasing the number of features can benefit the SVM results, all the methods were coordinated to have the same number of features. In this study, eight features plus the information about the locked excerpt were used.

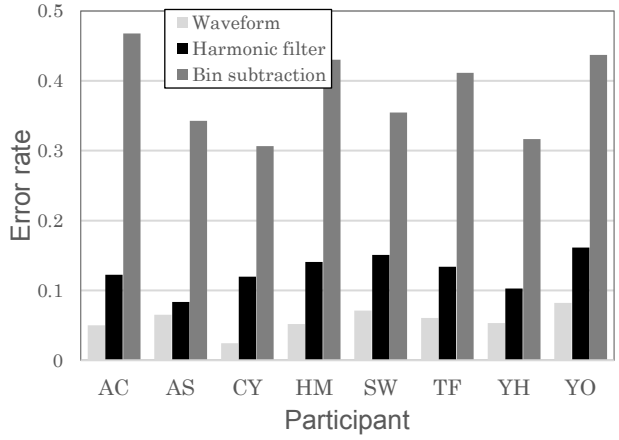
3. RESULTS

Data from two participants were removed: from one due to noise level contamination, and from another due to an incorrect number of answers for the given task.

The data obtained in the preliminary test was epoch-extracted and averaged for all participants. An example is shown in Fig 3. We can see that large EEG responses exist mainly in the frontal cortex. The responses differed with the type of modulation. We chose exponential modulation for the main test because it gave relatively stable responses compared to other modulations in many cases. Since we want to achieve our final goal using a simple EEG device, the number of channels has to be limited. Hereafter, the data shown were calculated only from the Fz channel. Although the results are not shown, we checked the other channels in the frontal area to confirm that they have similar performance.

For the main experiment, Fig. 4 shows the results for the SVM applied to the three types of feature extraction evaluated by ten-fold cross-validation. Figure 5 shows the subjective evaluation the ease of staying focused on the indicated excerpt and neglecting the other.

a)



b)

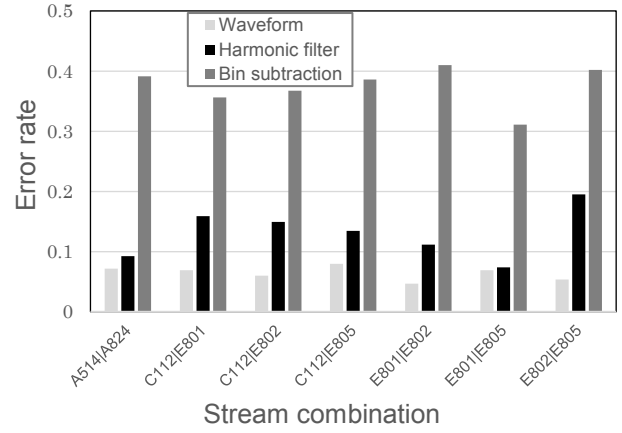


Fig. 4. Error rates of ten-fold cross-validation using SVM. SVMs were used for each combination of participants and stimuli. a) Error rate averaged over stimuli to see the difference between participants. b) Error rate averaged over participant to see the difference between stimuli.

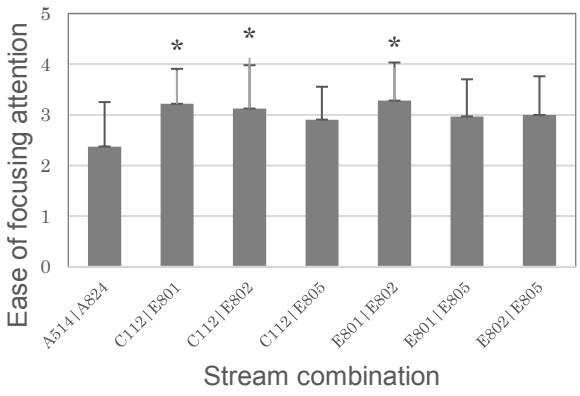


Fig. 5. Averaged evaluation of ease of focusing attention on the indicated stream. A significant difference was found with ANOVA $F(6)=3.47$, $p<0.01$. Subjective tests revealed significances only for the pair between A514A824, which are suggested with * for $p<0.05$.

4. DISCUSSION

Although musical excerpts are expected to have a great advantage in BCIs, few studies have used them for BCI purposes. From Fig. 5, we can confirm that even modulated musical excerpts are better than artificial stimuli for keeping focus on the stream. In Fig. 3, we can see prominent responses locked to the beat onsets and that the proposed exponential modulation gives the steepest and fastest EEG responses. The responses are mainly observed in the frontal area, is consistent with previous reports [3, 9].

We tested three methods for feature extraction. All methods used almost the same length of EEG data (around 10 s) to have the same number of features (eight features). Therefore, the error rates in Fig. 4 are considered to reflect the quality of the feature extractions. The bin-subtraction method has shown promising performance [9]. However, the averaged waveform method outperformed it and the harmonic filter method (Fig. 4). The reason may be the length used for calculation, but, more importantly, this could be due to fluctuations in the beat frequency. Since lower frequencies are vulnerable to time shifting, adaptive phase locking becomes important. In other words, our method succeeded in exploiting the ASSRs by precisely following the fluctuations both in modulation and in epoch extraction.

5. CONCLUSION

In this study, we showed that musical excerpts can be used for BCI purposes. To exploit the beat rhythm, we conducted modulation to emphasize the EEG response and used beat-locked feature extraction. Thanks to the precise beat locking, the error rate is low enough for a mass-EEG measurement system.

12. REFERENCES

[1] Jan. B. F. van Erp, F. Lotte, M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer Society*, 45(4), pp. 26–34, 2012. https://doi.org/10.1109/MC.2012.107

[2] T. G. Sato, Y. Shiraki, and T. Moriya, "Audience excitement reflected in respiratory phase synchronization," *IEEE Int. Conf. SMC*, pp. 2856–2860 2017.

[3] N. J. Hill, and B. Schölkopf, "An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli," *J. Neural Engineering*, 9, 026011, 2012, https://doi.org/10.1088/1741-2560/9/2/026011.

[4] Y. Mahajan, C. Davis, and J. Kim, "Attentional modulation of auditory steady-state responses," *PLoS ONE*, 9(10), 2014. https://doi.org/10.1371/journal.pone.0110902.

[5] D.W. Kim, H. J. Hwang, J. H. Lim, Y.H. Lee, K. Y. Jung, and C. H. Im, "Classification of selective attention to auditory stimuli: Toward vision-free brain-computer interfacing," *J. Neuroscience Methods*, 197 (1), pp. 180–185, 2011, https://doi.org/10.1016/j.jneumeth.2011.02.007.

[6] S. Kanoh, K. I. Miyamoto, and T. Yoshinobu, "A brain-computer interface (BCI) system based on auditory stream segregation," *J. Biomechanical Science and Engineering*, 5(1), pp. 32–40, 2010, https://doi.org/10.1299/jbse.5.32.

[7] I. Nambu, M. Ebisawa, M. Kogure, S. Yano, H. Hokari, and Y. Wada, "Estimating the Intended Sound Direction of the User: Toward an Auditory Brain-Computer Interface Using Out-of-Head Sound Localization," *PLoS ONE*, 8(2), 2013, https://doi.org/10.1371/journal.pone.0057174

[8] C. Pokorny, D. S. Klobassa, G. Pichler, H. Erlbeck, R.G. L. Real, A. Kübler, G. R. Müller-Putz, "The auditory P300-based single-switch brain-computer interface: Paradigm transition from healthy subjects to minimally conscious patients," *Artificial Intelligence in Medicine*, 59(2), pp. 81–90, 2013, https://doi.org/10.1016/j.artmed.2013.07.003

[9] S. Nozaradan, I. Peretz, M. Missal, and A. Mouraux, "Tagging the Neuronal Entrainment to Beat and Meter," J. Neuroscience, 31(28), pp. 10234–10240, 2011, https://doi.org/10.1523/JNEUROSCI.0411-11.2011

[10] H. Okawa, K. Suefusa, T. Tanaka, "Neural Entrainment to Auditory Imagery of Rhythms," Frontiers in Human Neuroscience, 11, pp. 1–11, 2017, https://doi.org/10.3389/fnhum.2017.00493

[11] T. Lenc, P. E. Keller, M. Varlet, and S. Nozaradan, "Neural tracking of the musical beat is enhanced by low-frequency sounds," Proceedings of the National Academy of Sciences, 201801421, 2018, https://doi.org/10.1073/pnas.1801421115

[12] R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain, "Name that tune: Decoding music from the listening brain," NeuroImage, 56(2), pp. 843–849, 2011, https://doi.org/10.1016/j.neuroimage.2010.05.084

[13] A Delorme and S Makeig "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," J. Neuroscience Methods, 134, pp. 9-21, 2004.

[14] B. E. Boser, M. G. Isabelle, and N. V. Vladimir, "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992.