Accepted Manuscript

Title: A novel hybrid auditory BCI paradigm combining ASSR and P300

Authors: Netiwit Kaongoen, Sungho Jo

PII:	S0165-0270(17)30017-1
DOI:	http://dx.doi.org/doi:10.1016/j.jneumeth.2017.01.011
Reference:	NSM 7666
To appear in:	*Journal of Neuroscience Methods*
Received date:	16-11-2016
Accepted date:	14-1-2017

Please cite this article as: Kaongoen Netiwit, Jo Sungho.A novel hybrid auditory BCI paradigm combining ASSR and P300.*Journal of Neuroscience Methods* http://dx.doi.org/10.1016/j.jneumeth.2017.01.011

A Novel Hybrid Auditory BCI Paradigm Combining ASSR and P300

Netiwit Kaongoen[a], Sungho Jo[a,*]

[a]Intelligent System and Neurobotics Lab (ISNL), School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

E-mail address: netiwit.k@cs.kaist.ac.kr (N. Kaongoen), shjo@kaist.ac.kr (S. Jo).*Corresponding author. Tel.: +82 42 350 7740.

Highlights

- A hybrid auditory BCI paradigm combining ASSR and P300 BCI is proposed.
- The distinct AM frequency of each sound source triggers the ASSR and increases the difference between the sound sources in spatial-auditory P300 BCI system.
- Combining ASSR and P300 BCI into a hybrid system results in a better performance.

Abstract

Background: Brain-computer interface (BCI) is a technology that provides an alternative way of communication by translating brain activities into digital commands. Due to the incapability of using the vision-dependent BCI for patients who have visual impairment, auditory stimuli have been used to substitute the conventional visual stimuli.

New Method: This paper introduces a hybrid auditory BCI that utilizes and combines auditory steady state response (ASSR) and spatial-auditory P300 BCI to improve the performance for the auditory BCI system. The system works by simultaneously presenting auditory stimuli with different pitches and amplitude modulation (AM) frequencies to the user with beep sounds occurring randomly between all sound sources. Attention to different auditory stimuli yields different ASSR and beep sounds trigger the P300 response when they occur in the target channel, thus the system can utilize both features for classification.

Results: The proposed ASSR/P300-hybrid auditory BCI system achieves 85.33% accuracy with 9.11 bits/min information transfer rate (ITR) in binary classification problem.

Comparison with Existing Methods: The proposed system outperformed the P300 BCI system (74.58% accuracy with 4.18 bits/min ITR) and the ASSR BCI system (66.68% accuracy with 2.01 bits/min ITR) in binary-class problem. The system is completely vision-independent.

Conclusions: This work demonstrates that combining ASSR and P300 BCI into a hybrid system could result in a better performance and could help in the development of the future auditory BCI.

Keywords Brain computer interface; Auditory P300; Auditory steady state response; Hybrid system.

1. Introduction

Brain-computer interface (BCI) is a technology that provides an alternative way of communication without using any kind of muscular activity by measuring and translating brain activities into digital commands that can be transformed into messages or used to control electrical devices. Initially, BCI technology was invented to help patients who have severe motor impairment such as those who suffer from serious injury or neurological diseases (e.g.. amyotrophic lateral sclerosis (ALS) disease) [1]. BCI systems can be categorized into 3 types: active, passive and reactive [2], based on mental task that elicits distinct brain patterns and experiment paradigm. In active BCI, users consciously perform mental task such as motor imagery [3], and mental calculation [4]. Passive BCI derives its outputs from arbitrary brain activities without the voluntary control, e.g. mental state and emotion detection [5]. The last type is reactive BCI which utilizes the brain activity response to external stimuli. Reactive BCI is relatively easy and requires less effort since user can just observe the given stimuli and let the system do all the work by detecting and utilizing the natural brain activity responses. The examples are steady state visually evoked potential (SSVEP) and P300 BCI. SSVEP is the stable oscillation in voltage of brain electrical activity and it has same or resonant frequency as the visual stimuli. P300 is a positive deflection that occurs in the electroencephalogram (EEG) approximately 300 ms (its latency may vary from 250 to 750 ms) after a stimulus that is delivered under a specific set of circumstances known as the Oddball paradigm. A series of events (i.e., stimuli) that fall into one of the two classes (target and nontarget) will be presented to the subject by having the target stimuli less frequent than nontarget stimuli [6].

Studies have shown that combining two BCIs, or at least one BCI and another system into a hybrid system is another way to improve the performance [7]. The hybrid BCI can either have multiple inputs that are typically processed simultaneously or operate two systems sequentially in which the first system acts as the brain switch. Hybrid BCI system in [8] incorporated the SSVEP into P300 speller system. They designed a periodic stimuli mechanism that works as the trigger to the SSVEP response and superimpose it onto the P300 stimuli to increase the difference between the characters in the same row or column. This proposed hybrid BCI speller system was shown to achieve a better and more stable performance than the conventional P300 speller.

Most of the reactive BCI techniques that have been invented so far rely on a visual stimulation or involve eye-gazing to the stimuli. In reality, these kinds of vision-dependent BCI techniques such as SSVEP and conventional visual-based P300 methods are not be able to provide reliable communication channel to users who have visual impairment or patients in completely lock-in state (CLS) that lost the control in their extraocular movement. To overcome this problem of vision-dependent BCI paradigms, researcher have replaced visual stimuli with other sensory stimuli including auditory stimuli [9][10]. In [9], an auditory version of P300 speller was introduced and compared with the visual modality. This auditory P300 speller uses acoustically presented numbers to code each number in the speller matrix instead of flashes in row and column of the matrix in the conventional visual P300 speller. The results showed that auditory-based P300 speller is feasible but the performance was inferior to the visual-based system. The work in [11] goes with another approach to construct the auditory speller system. They use two-dimensional auditory stimuli to code the speller matrix: variation in pitch and location of the auditory stimuli represent each row and column of the speller matrix, respectively. Many studies introduced and performed experiments with different kinds of auditory stimuli and experiment paradigms to improve the performance of auditory P300 BCI. For instance, [12] compared the effect of three physical differences of auditory stimuli: pitch, loudness and direction, in the binary auditory P300 BCI. Work in [13] presented auditory multi-class BCI paradigm using spatial hearing (i.e., not only consider which sounds to attend but also consider the location of the sound source) as an informative cue in the Oddball paradigm. In their system, auditory stimuli were randomly presented from five speakers surrounding the subject in semi-circle form with same distance from the subject and were spaced equally between each other. In their experiment, subjects have to focus on one of the five speakers and then P300 responses can be detected and interpreted. The system was successful with high accuracy and information transfer rate for the healthy subjects and thus, spatial-auditory based P300 is a promising system for auditory based BCI.

As an alternative to auditory-based P300 BCI, study in [14] conducted experiments to test the possibility of using auditory steady-state response (ASSR) as a new BCI paradigm. ASSR is an electro-physiologic auditory evoked potential (AEP) that responds to the amplitude-modulation (AM) frequency of the auditory stimulus by oscillating with the same or resonant frequency. ASSR exhibits increment in spectral density around AM frequency of the sound signal. This research provides the evidence that selective attention can modulate ASSR and thus, demonstrated the possibility of using ASSR modulated by auditory selective attention as a new auditory BCI paradigm. Following the work in [14], study in [15] investigates the feasibility of using ASSR in binary-class auditory BCI by using two speakers placing in front of the subject with same distance. Each speaker simultaneously plays a sound with distinct pitch and AM frequency. Subjects have to selectively focus on one speaker according to the given instruction and EEG signal is acquired. The power spectral density (PSD) values for each AM frequencies of the sound sources are used as features in the classification model. The results showed that the classification accuracy is higher when the length of stimuli is longer. In addition, recent work [16] conducted the ASSR BCI experiment in the similar setting to [15] but used spatial coherence to detect the channel that the subject is focusing, which showed similar results to [15].

ASSR is relatively new to the BCI community. Unlike P300 BCI ASSR BCI was only examined in the binary classification problem and have not yet shown as a reliable system to be used in real life application. The objective of this study is, therefore, to serve as a preliminary study that introduces the possibility of combining ASSR BCI system and spatial-auditory BCI system into a hybrid BCI system to improve the performance of the auditory BCI. In the proposed system, sounds with different pitches and AM frequencies are simultaneously presented to the subject separately through different sound channel. Each sound channel is then separately increased in volume for short amount of time (resulting in 'beep' sound) randomly according to the Oddball paradigm. With this setting, both ASSR and spatial-auditory P300 features can be detected and used for the further analysis and classification. The proposed BCI paradigm is analogous to P300/SSVEP hybrid BCI method [9]. We test the proposed hybrid system in the binary classification problem for simplicity. We hypothesize that the ASSR/P300 hybrid system will have better performance than the BCI system in sole ASSR and spatial-auditory P300 condition. The system will be described in details in the following sections.

2. Method and experiment

The ASSR/P300 hybrid auditory BCI works by combining ASSR and P300 stimulus together to form a new hybrid stimulus. The auditory stimulus from each sound source will have distinct pitch and AM frequency that triggers the ASSR response while also have beep sounds occurring randomly in each sound channel according to the Oddball paradigm that will trigger the P300 responses. EEG signal recorded from the subject while attending to each stimulus is then processed separately in ASSR and P300 computing module and, the prediction results from both modules are fused together with the fusion method to give the final prediction results. The overview of the proposed system is shown in Figure 1. The details of the methods and experiments are provided in the following sections.

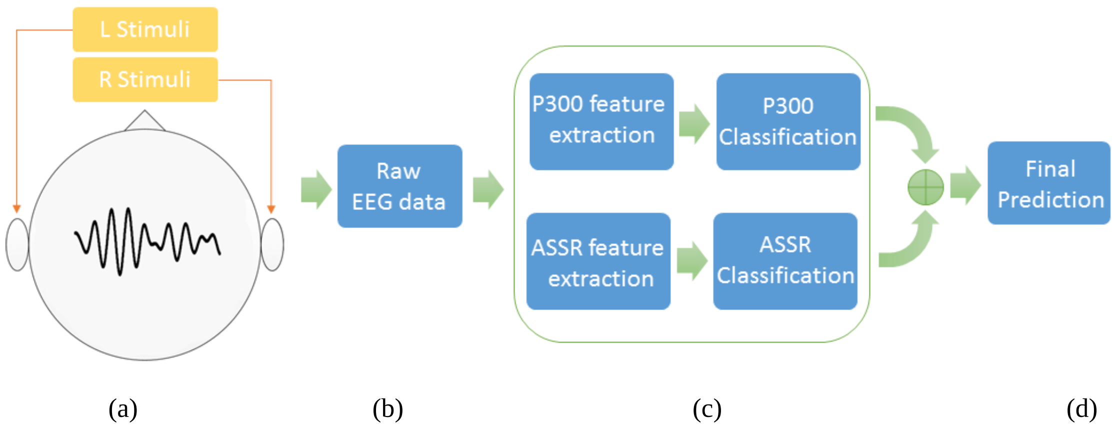

Figure 1. The overview of ASSR/P300 hybrid auditory BCI system. (a) Two different auditory stimuli were given to the subject separately in left and right channel. (b) Raw EEG data acquisition. (c) Raw data are processed and the feature vector are formed to be used in the classification method. (d) The prediction results from P300 and ASSR module are combined by the fusion method to give the final prediction results.

2.1) ASSR/P300-hybrid auditory stimuli

Two sound sources with different pitch and AM frequency were used as the auditory stimuli in our proposed ASSR/P300-hybrid auditory BCI system. In this experiment, two sounds were set to have 1 kHz and 2 kHz pitch with 37 Hz and 43 Hz AM frequency, respectively. Both sounds were generated by using MATLAB at a sampling rate of 44,100 Hz in the waveform audio file format (*.wav). Both sounds were simultaneously presented separately in the left and right sound channels of the subject: the 1 kHz sound in the left channel and 2 kHz sound in the right channel. In one block of stimuli, either left or right channel was chosen as the target channel and ten beep sounds (brief static increase in the volume) were presented to the subject. Beep sounds occured two times in the target channel and eight times in the nontarget channel. Each beep sound lasted for 100 ms following by 200 ms of base volume as the interstimulus interval (ISI) (time between the offset of a stimulus to the onset of another one) of P300 stimuli, resulting in the total length of 3 seconds for each block of stimuli. An example of a block of stimuli is depicted in Figure 3. These beep sounds trigger P300 responses and selectively attending to left or right auditory stimulus yields different ASSR from the subject. Therefore, by using this hybrid stimuli, both ASSR and P300 feature can be extracted from the raw EEG data and used for further analysis and classification.

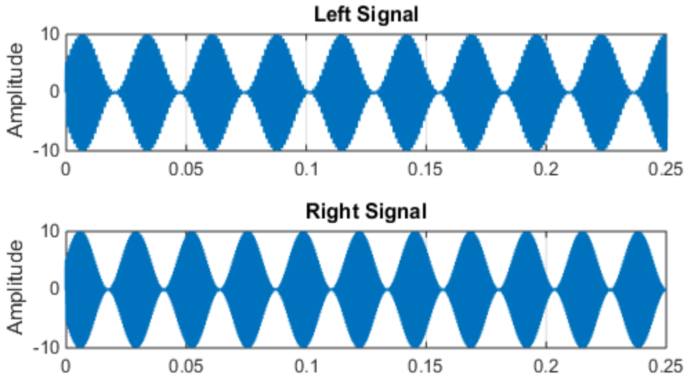

Figure 2. The sound stimuli used in the experiment: 1.0 kHz sound with 37 Hz AM frequency (Left channel) and 2.0 kHz sound with 43 Hz AM frequency (Right channel)

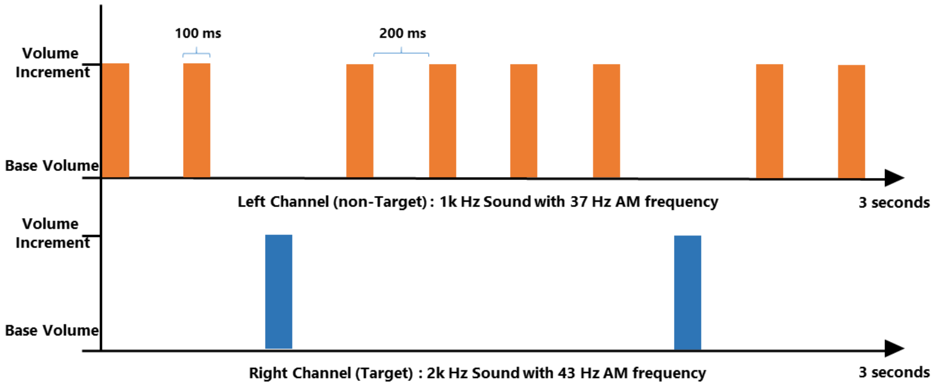

Figure 3. An example of stimuli in one block having left channel as the nontarget and right channel as the target. The beep sound lasts 100 ms and the ISI is equal to 200 ms

2.2) Participants

Ten healthy undergraduate students from Korea Advance Institute of Science and Technology (KAIST) voluntarily participated as the subjects in our experiment. Six of the subjects are male (aged 20±1 years) and four of the subjects are female (aged 20±2 years). All subjects gave written informed consents. The KAIST Institutional Review Board approved the proposed experimental protocol of this study. All of the subjects were free from any neurological disorders, visual and hearing impairment, and had never experienced any kind of auditory BCI experiment.

2.3) Experiment setup

Before the experiment begins, the experiment procedures were explained to the subject via both written and verbal instruction. Each subject was presented with left and right auditory stimulus to learn and memorize the auditory stimulus of each channel and was also asked to adjust the base volume to the volume that they are comfortable with. Subjects were asked to sit comfortably and put on an earphone (MDRXB50AP, SONY) with eyes closed. One run of the experiment consists of two trials, one with the left channel as the target and another one with the right channel as the target. Each run starts with 5 seconds of both sounds playing simultaneously separately in each sound channel at the base volume (labeled as 'rest time') followed by the first trial (having left channel as the target). One trial consists of seven blocks of stimuli in which the subjects have to attend and count whenever the beep sound occurs in the target channel. Two trials are separated with 5 seconds of rest time and the experiment ends with another 5 seconds of rest time. The total running time for one run of the experiment was 57 seconds. To make sure that subjects really understand the experiment procedure, one additional run of the experiment was performed before the first run begins. One subject was subjected to the experiment in the total of two sessions, each consists of ten runs, with 5 minutes interval between sessions. All subjects were also asked to give some feedback or report any trouble during/after the experiment. The experiment took approximately 45 minutes per one subject.

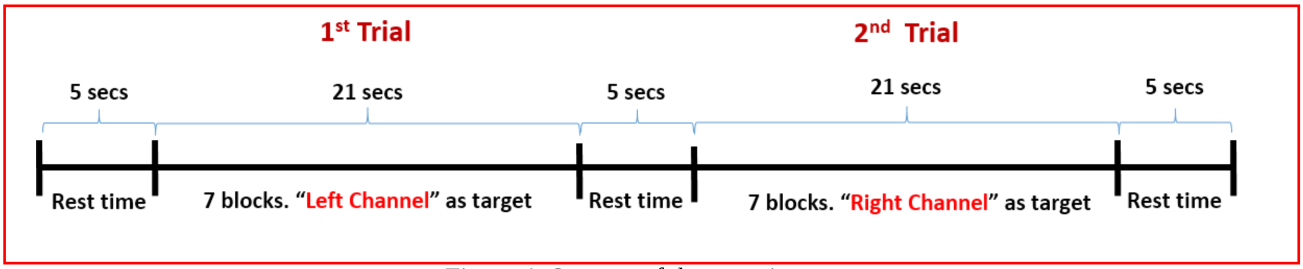

Figure 4. One run of the experiment

2.4) Data acquisition

EEG data was recorded using OpenBCI 32bit board kit (www.openBCI.com) with passive gold cup electrodes and EEG conductive paste in eight channels including Fz, Cz, Pz, P3, P4, Oz, T3 and T4 according to the International 10-20 system (Figure 5). Six channels including Fz, Cs, Pz, P3, P4, and Oz were used for P300 analysis and six channels including Fz, Cz, Pz, Oz, T3 and T4 were used for ASSR analysis. All eight EEG channels were referenced to the right earlobe and grounded to the left earlobe. The sampling rate of EEG was

250 Hz. After acquiring the EEG signal from each trial of the experiment, EEG data are segmented into EEG epochs for both P300 and ASSR analysis and underwent processing steps to construct the feature vector to be used later in the classification process. In P300 analysis, EEG epoch with lengths of 800 ms from the stimulus onset is cut to represent the brain activity response to each P300 stimulus. It should be noted that in each EEG epoch, the last 600 ms of data overlap with the consecutive EEG epoch. For ASSR analysis, the entire 3 seconds of EEG is cut to represent the ASSR acquired in each block of stimuli.

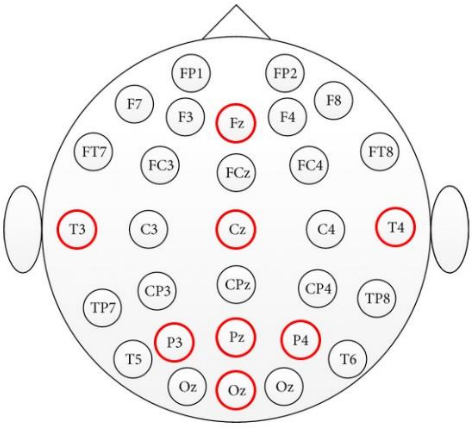

Figure 5. The International 10-20 system for EEG node positions. The total of eight channels are used in this work including Fz, Cz, Pz, P3, P4, Oz, T3 and T4 (labelled with red circles).

2.5) Data preprocessing

In our system, EEG data for ASSR and P300 analysis undergo the processing method separately to construct the feature vector for the classification process. ASSR EEG epochs are filtered using 5th-order 0.1 Hz – 50.0 Hz Butterworth bandpass filter. Canonical Correlation Analysis (CCA) is chosen as the ASSR feature extraction method. CCA explores the correlation between EEG signal and reference signals. It has been widely used in SSVEP-based BCIs [17, 18]. In our case, the reference signals are defined as the set of sin-cos signals with the frequency identical and resonant to the ASSR stimuli as follow:

$$y_f(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \\ y_5(t) \\ y_6(t) \end{pmatrix} = \begin{pmatrix} \sin(2\pi f t) \\ \cos(2\pi f t) \\ \sin(4\pi f t) \\ \cos(4\pi f t) \\ \sin(6\pi f t) \\ \cos(6\pi f t) \end{pmatrix}, \quad t = \frac{1}{s}, \frac{2}{s}, \dots, \frac{\mathrm{T}}{s}, \quad f = 37Hz, 43Hz \tag{1}$$

where f is the AM frequency of the ASSR stimuli, T is the number of data point, and s is the sampling rate.

CCA is applied to the multi-channel ASSR EEG epochs from each trial and both sets of the reference signals. The output correlation values from both set of reference signals are then used as the feature for the classification in ASSR module.

In P300 case, EEG data are filtered with 5^{th}-order Butterworth bandpass filter with 1.0 Hz – 12.0 Hz cutoff frequency. Winsorization is then performed to remove the effects of outliers in EEG data for better signal quality. In this system, the fifth and ninety-fifth percentile were computed for each EEG channel and the values lower or higher than these percentiles are replaced by the fifth and ninety-fifth percentile values, respectively. EEG data for P300 analysis are then segmented into epochs as explained in the section 2.4 and standardized to have zero means and standard deviations equal to one. Each EEG epochs are decimated with a factor of 15: decimation process applies an 8^{th}-order lowpass Chebyshev Type I IIR filter to guard against aliasing of the signal and down-samples the signal by keeping only every 15^{th} sample of the data. The preprocessed EEG epochs have the length of 13 data point. Finally, the feature vector are formed by concatenating the preprocessed EEG epochs from all six channels and labelled with -1 for nontarget and 1 for target response.

2.6) Classification and evaluation

2.6.1) Classification

The proposed system was evaluated with 10-fold cross-validation method. In each iteration of the cross-validation, data from eighteen runs were used as training data while another two runs were used as testing data. Both P300 and ASSR classification model were trained separately using Fisher's Linear Discriminant Analysis (FLDA).

Each trial is classified in 2 cases: single-block and multiple-block. In single-block case, the system uses only EEG data obtained from only the first block of stimuli of each trial to analyze and classify the trial. In multiple-block case, the prediction results obtained from each block are averaged to get the final prediction results for a trial.

For ASSR, the classification for each trial is done by directly using the feature vector constructed from ASSR EEG epoch. In P300 case, all P300 EEG epochs in a trial are first classified using the model trained from the training dataset to get the prediction score whether each P300 EEG epoch is P300 response or not. To classify each trial, we determine the average of the prediction scores for the target and nontarget EEG epochs separately as followed:

$$Score^{P300}_{non-Target} = Mean(Score^{P300}_{i}), \qquad i \in \text{[nontarget indices]} \tag{2}$$

$$Score^{P300}_{Target} = Mean(Score^{P300}_{i}), \qquad i \in \text{[Target indices]} \tag{3}$$

With this method, we the trial is classified correctly (subject has P300 response to the target stimuli) if and only if $Score^{P300}_{Target}$ is higher than $Score^{P300}_{non-Target}$.

It should be noticed that the classification in ASSR deals with the problem whether the subject is focusing on left or right signal while the classification in P300 deals with the problem whether the subject is attending to the target channel or not.

2.6.2) Fusion method

Fusion methods are categorized into two main types [20]. The first method is based on heuristic rules product such as max, min, average and voting rules. The second method is based on the probability such as the Bayes classifier. Since the second category of fusion method relies on large data sample, hence it is unsuitable for our study, and we decide to use weighted linear combination as the fusion method in our system.

In the proposed system, the fusion score is defined as:

$$Score^{Fusion}_{c} = w_{c1}Score^{ASSR}_{c} + w_{c2}Score^{P300}_{c}, \qquad c \in [L, R] \tag{5}$$

where w_{c1} and w_{c2} is the weight for prediction result from ASSR and P300 module, respectively. In P300 case, $Score^{P300}_{L}$ is $Score^{P300}_{Target}$ and $Score^{P300}_{R}$ is $Score^{P300}_{non-Target}$ in the trials that have left channel as the target channel and $Score^{P300}_{R}$ is $Score^{P300}_{Target}$ and $Score^{P300}_{L}$ is $Score^{P300}_{non-Target}$ in the trials that have right channel as the target channel.

We then define a decision hyperplane:

$$g(x) = w.x + b$$

where w is the vector of weight $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, x is the feature vector defined as $x = \begin{bmatrix} Score^{ASSR} \\ Score^{P300} \end{bmatrix}$, and b is the bias term. We treat this as another classification problem and use stepwise linear discriminants analysis (SWLDA) as the meta-classifier to find the optimal weight. This process is performed using the same training and testing indices from the 10-fold cross-validation in the classification step. The optimal weight is learned using the prediction scores from training set and then applied to the scores of the testing set to calculate the fusion scores. Finally, each trial is assigned to the class that have highest value of fusion score.

2.6.3) Evaluation

We evaluated our proposed system by computing both the accuracy of the classification and the information transfer rate (ITR). The accuracy is measured by the number of correctly-classified test trials from each iteration of the cross validation method. ITR measures how much information (in bit) the system can transfer in one minute [19] and it has been widely used in BCI community to measure the performance of the system. ITR is defined as:

$$\boxed{ITR = \left\{\log_2 N + Plog_2 P + (1-P)\log_2 \frac{1-P}{N-1}\right\}/T} \tag{4}$$

where N is the number of class, P is the classification accuracy, and T is the time interval for each selection in minute resulting the unit for ITR to be bits/min. The system is then evaluated in three manners: (1) Using only prediction results from ASSR classification module. (2) Using only prediction outputs from P300 classification module and (3) Applying the fusion method to combine the results from both P300 and ASSR modules to get the final prediction scores. Finally, since cross-validation method gives different results in every run, the evaluation method was repeated ten times and the averaged values for both accuracy and ITR were used as the final performance of the system.

3. Results

To verify that our subjects truly respond to the P300 and ASSR stimuli, EEG signals were averaged and plotted to compare when, in P300 case, the subject was presented with nontarget and target stimuli, and, in ASSR case, the subject attended to the 37 Hz stimulus and 43 Hz stimulus. Figure 6 (left) demonstrates filtered EEG data from the subject S1 showing the clear P300 in the blue plot occurred approximately 400 ms from the onset of the stimuli. Figure 6 (right) shows the power spectrum of EEG signal averaged across the subjects. We can observe peaks at the frequency around 37 Hz and 43 Hz. Moreover, we can see that the EEG signal when subjects attended to 37 Hz stimuli (red plot) has higher peak in 37 Hz and lower peak in 43 Hz comparing to the EEG signal when subjects attend to the 43 Hz stimuli (blue plot). This shows that the selective attention can modulate the ASSR which is correspondent with the previous studies [12-14].

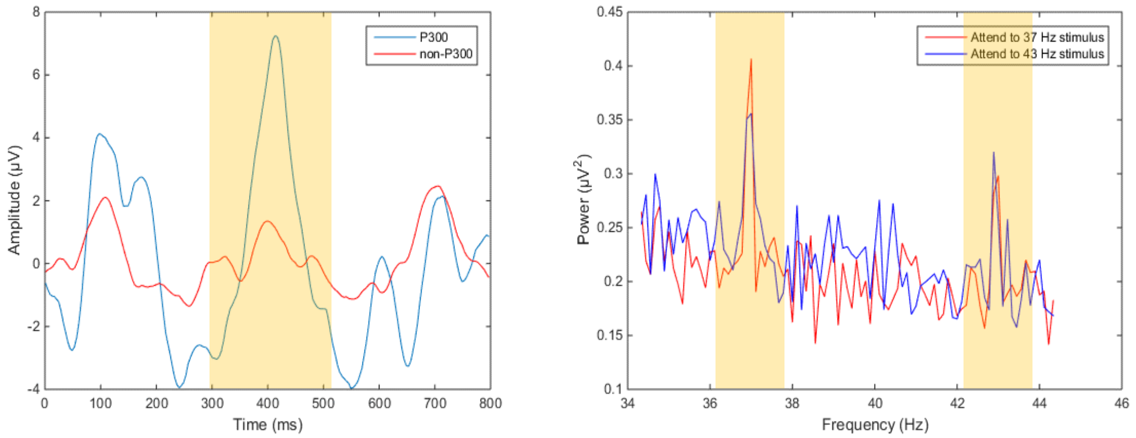

Figure 6. (left) P300 occurred approximately 400 ms after the onset of the stimuli. (right) Power spectrum of EEG signal averaged across the subjects. Peaks are higher in 37 Hz and lower in 43 Hz when subject selectively attend to the 37 Hz stimuli comparing to the 43 Hz one, and vice versa.

The averaged accuracy and ITR of each subject from single-block trials (20 commands/minute) is shown in the Table 1. From the results, we can see that the accuracy and ITR from the hybrid system is higher than both sole P300 and ASSR condition in all subjects. The accuracy was improved by 14.14% from the sole P300 condition ($\rho = 0.001$) and 27.97% from the sole ASSR condition ($\rho < 0.0001$) according to the ANOVA test when using the hybrid system. These results show that the proposed system is significantly better than the both of sole P300 and ASSR condition.

Table 1:

	Accuracy comparison (%)			ITR comparison (bits/min)		
Subject	Hybrid	P300	ASSR	Hybrid	P300	ASSR
S1	81.50	70.75	70.50	6.49	2.63	2.64
S2	96.50	84.25	72.50	16.48	7.51	3.26
S3	90.00	75.50	72.00	10.81	4.19	3.08
S4	76.25	66.75	61.75	4.68	1.79	1.23
S5	69.25	61.25	60.25	2.40	1.02	0.94
S6	79.50	69.50	63.75	5.36	2.25	1.11
S7	93.75	84.25	56.25	13.25	7.43	0.23
S8	83.25	74.00	69.00	6.96	3.47	2.14
S9	88.00	75.00	70.25	9.80	3.88	2.58
S10	95.25	84.50	70.50	14.82	7.62	2.84
AVG	**85.33**	**74.58**	**66.68**	**9.11**	**4.18**	**2.01**
STD	8.47	7.53	5.41	4.45	2.37	2.13

Comparison in accuracy and ITR using single block of stimuli (20 commands/minute)

To illustrate the feasibility of our hybrid system, the accuracy and ITR in multiple-block case were also computed for all subjects. The accuracy in all settings increases as we used more blocks of stimuli in a trial. In all subjects, the hybrid system gives the best accuracy following by the sole P300 and ASSR condition. By using the hybrid system, five out of ten subjects were able to reach 100% accuracy outnumbering the sole P300 condition in which only two out of ten subject were able to reach 100% accuracy when we increase the number of block used in a trial. Even though the accuracy increases as we increase the number of block used in a trial, it would take longer time for the system to execute an output command, so most of the subject has highest ITR when using prediction results from single-block. The averaged accuracy and ITR result across the subjects are shown in Table 1 and Table 2, respectively. In descending order, the maximum averaged accuracy for each setting are hybrid system: 99.28% (seven blocks), P300: 98.38% (seven blocks), and ASSR: 85.42% (six blocks). The maximum averaged ITR for each setting in descending order are hybrid system: 9.105 bits/min, P300: 4.181 bits/min, and ASSR: 2.005 bits/min. Maximum ITR of all three settings is from the single-block trial.

Table 2: The averaged accuracy (%) across the subjects between all settings

Setting	**Number of block**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
P300	74.58	84.82	91.22	94.65	96.43	97.50	**98.38**
ASSR	66.67	72.70	79.45	81.90	85.22	**85.42**	82.90
Hybrid	85.32	90.30	95.05	96.55	97.88	98.98	**99.28**

Table 3: The averaged ITR (bits/min) across the subjects between all settings

Setting	**Number of block**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
P300	**4.181**	4.141	3.882	3.595	3.202	2.837	2.566
ASSR	**2.005**	1.734	1.987	1.697	1.661	1.445	1.028
Hybrid	**9.105**	5.862	4.923	4.033	3.496	3.125	2.731

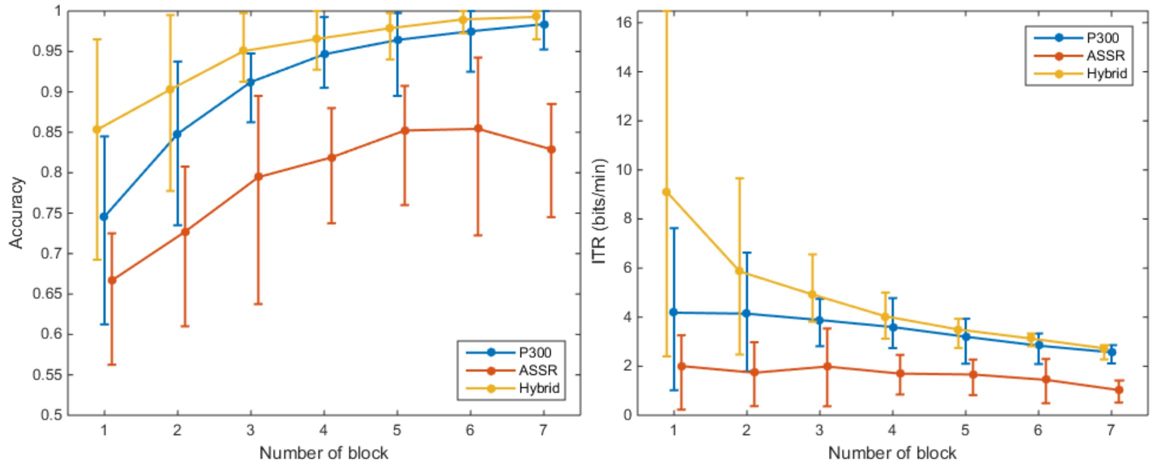

Figure 7. Comparison of the averaged (left) accuracy and (right) ITR between all settings. Error bars indicate the maximum and minimum values across the subjects.

4. Discussion

In this study, we presented an auditory hybrid BCI system with the goal of improving the performance of auditory BCI. Inspired by SSVEP/P300-hybrid BCI speller paradigm that each character in the same row or column of the speller matrix flickers with different frequencies to increase the difference between characters, our hybrid system presented subjects with spatially separated auditory stimuli with different pitches and AM frequencies with beep sounds occurring randomly in each channel. The proposed system was able to induce both ASSR and P300 responses in subjects' EEG signal. Based on the results performing on ten healthy subjects, the hybrid system had better performance than the ASSR and P300 BCI individually in the binary-class problem, demonstrating the feasibility of incorporating ASSR and spatial auditory P300 BCI into an auditory hybrid system.

Our hybrid auditory BCI paradigm is completely vision-free and suitable to be used for the users who has a visual impairment and cannot use the conventional vision-dependent BCI systems. The system is practical and simple. Subjects could easily understand how the system works. In fact, no subject reports any kind of trouble in performing the experiments that they have to concentrate and count whenever beep sound occurs in the target channel. Another advantage of hybrid BCI system is that it could help solving the problem of BCI illiteracy that some people might have difficulty to use some specific BCI system due to their uniqueness in the brain activity [21]. Regardless of our results, some users might have better performance when using the P300 BCI system than the ASSR BCI system while some might show the opposite result. By using a hybrid method, the system can be trained and learnt the best configuration (i.e., weights in fusion method) for the system so that it fits to each individual, resulting in the best outcome.

Despite the advantages of the proposed hybrid auditory BCI system discussed above, the present ASSR/P300-hybrid auditory BCI paradigm needs to be adjusted and refined before using in the real-life application. Studies have shown that humans have strongest ASSR with the 40 Hz stimuli but the reason is still remained unclear [14]. In this study, we used frequencies around 40 Hz: 37 Hz and 43 Hz for all subjects as the choices for binary-class BCI, however, each individual might have their own optimal frequencies of ASSR stimuli, thus one should consider selecting the optimized set of ASSR stimuli for a user to be used in practical application. There are also many variables that could be adjusted to optimize the system. The ISI could be reduced so that the command can be sent faster. The effect of beep sounds (P300 stimuli) to the ASSR is unclear in the current study. Beep sounds might disturb the subject’s attention on ASSR stimuli affecting the performance of ASSR BCI module. Lowering the volume of beep sound might help alleviate the disturbance, but it might also degrade the P300 responses, therefore, further studies are needed to determine this tradeoff.

Although this work has shown the supremacy of the proposed hybrid paradigm over the conventional auditory P300 BCI in the binary class problem, it is still unclear whether the proposed hybrid system will be a better approach for auditory BCI in the state-of-the-art. Thus, the future work would be to extend the current hybrid paradigm to multiple-class BCI system. We strongly believe that this method can as well be applied to the multiple-class paradigm. The simplest way to do this is by increasing the number of sound sources that each simultaneously plays a sound with different pitch and AM frequency to represent more classes. However, there have not been any report on the multiple-class ASSR-based BCI paradigm and thus, it remains questionable. Also, since our system simultaneously presents all auditory stimuli to the subject, it might be too difficult for the subjects to distinguish the target auditory stimulus from nontarget ones. Nevertheless, incorporating binary-class ASSR paradigm into multiple-class auditory P300-based BCI is absolutely possible. Taking the multiple-class auditory P300-based BCI system described in [13] as an example, eight speakers could be divided into two groups playing sounds with 37 Hz and 43 Hz AM frequency, respectively. With this setup, in the scenario where the prediction output from P300 model of two targets from the different group are similar and both targets could be the choice which the user selects, the system can use the prediction scores from ASSR module as the additional information to classify the final output. We can also construct the multiple-class paradigm even if we limit the number of sound source to two so that the system can be used with an earphone. For instance, we can construct an eight-class system by setting the P300 stimulus (e.g., beep sound) to four pitches (e.g., 1 kHz, 1.5 kHz, 2.0 kHz, and 2.5 kHz) and two channels that each plays a control sound with different AM frequency. Although this method might bias P300 module over ASSR module but we are certain that this can improve the performance of the system. In addition, the musical background of each subject might be one of the factors that affects the performance of the system. Subjects with musical background might be more capable of distinguishing specific sound from multiple sound sources surrounding them. A study that compares the performance between a group of subjects with a musical background and a control group would be interesting.

This preliminary study has shown the possibility of combining ASSR and auditory P300 to improve the performance of auditory BCI. The proposed hybrid auditory BCI is possible for a wide range of applications. Target sounds can be mapped to any kind of choice. In communication, each target sound can be simply mapped to a predefined phase or word (e.g., "YES", "NO", "PASS", "END", etc.) [22]. The proposed method can also be applied to auditory BCI speller but probably only to the system which the speller matrix is coded with pure tone sounds like the system shown in [11]. The natural sounds and spoken words do not have steady amplitude and that possibly interferes with the ASSR even though we modulate its amplitude with a specific AM frequency. Thus, the proposed hybrid paradigm is unlikely to be applicable to those systems that use that kind of auditory cue such as BCI speller system in [9] and [10]. The pure auditory P300 speller system would require the users to memorize the mapping of stimuli to the letters. This might sounds difficult and impractical but it is possible especially once the training has been done extensively. The hybrid auditory BCI can be used to control electrical devices such as home appliances [23] and electrical wheelchair [24] when the target sounds are mapped to the commands. For instance, in the wheelchair control, focusing on low-pitch beep sound in the left earphone that play a control sound with 37 Hz AM frequency can execute "go straight" command, focusing on the same beep sound but in the right earphone that plays a control sound with 43 Hz can execute "go back" command, and focusing on high pitch in the left and right earphone could execute "turn left" and "turn right" command, respectively. Auditory instructions can be used to guide the user which sound corresponds to which choice at the beginning of the system or every time before the choice is made. The auditory feedback can also be given to report the state or the result after the system executes the selected command. In addition, for non-patient user, the auditory P300 BCI has an advantage over the visual P300 BCI in the sense that the user's eyes can engage with something else while using the interface. This will be very functional in some scenarios such as using a BCI while driving that requires the user to engage with the environment. We are certain that this preliminary research will be helpful for the BCI researches, especially in the development of auditory BCI.

5. Acknowledgement

This work was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education [2013R1A1A2009378].

References

[1] Wolpaw JR, Wolpow EW. Brain-computer Interfaces: Principles and Practice. Oxford: Oxford UP (2012) Print.

[2] Zander TO, Kothe C, Jatzev S, Gaertner M. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. Brain-computer interfaces. Springer London. (2010): 181-199.

[3] Chae Y, Jeong J, Jo S. Toward brain-actuated humanoid robots: asynchronous direct control using an EEG-based BCI. IEEE Transactions on Robotics 28.5 (2012): 1131-1144.

[4] Power SD, Kushki A, Chau T. Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI. BMC research notes 5.1 (2012): 1.

[5] Lin CT, Chen YC, Huang TY, Chiu TT. Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning. IEEE Transactions on Biomedical Engineering 55.5 (2008): 1582-1591.

[6] Choi B, Jo S. A low-cost EEG system-based hybrid brain-computer interface for humanoid robot navigation and recognition. PloS one 8.9 (2013): e74583.

[7] Pfurtscheller G, Allison BZ. The hybrid BCI. Frontiers in neuroscience 4 (2010): 3.

[8] Yin E, Zhou Z, Jiang J, Chen F, Liu Y. A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. Journal of neural engineering 10.2 (2013): 026012.

[9] Furdea A, Halder S, Krusienski DJ, Bross D. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology 46.3 (2009): 617-625.

[10] Klobassa DS, Vaughan TM, Brunner P, Schwartz NE, Wolpaw JR., Neuper C, Sellers EW. (2009). Toward a high-throughput auditory P300-based brain–computer interface. Clinical neurophysiology, 120(7), 1252-1261.

[11] Höhne J, Schreuder M, Blankertz B, Tangermann M. Two-dimensional auditory p300 speller with predictive text system. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (2010): 4185-4188.

[12] Halder S, Rea M, Andreoni R, Nijiboer F, Hammer EM, Kleih SC, Birbaumer N., Kubler A. An auditory oddball brain–computer interface for binary choices. Clinical Neurophysiology 121.4 (2010): 516-523.

[13] Schreuder M, Blankertz B, Tangermann M. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PloS one 5.4 (2010): e9813.

[14] Lopez MA, Pomares H, Pelayo F, Urquiza J, Perez J. Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain–computer interfaces. Neurocomputing 72.16 (2009): 3617-3623.

[15] Kim DW, Hwang HJ, Lim JH, Lee YH, Jung KY, Im CH. Classification of selective attention to auditory stimuli: toward vision-free brain–computer interfacing. Journal of neuroscience methods 197.1 (2011): 180-185.

[16] Felix LB, Netto AA, Ranaudo FS, Miranda AMFL. A spatial coherence-based vision-free brain-computer interface using auditory selective attention. MedicalExpress 2.6 (2015).

[17] Lin, Zhonglin, et al. "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs." IEEE Transactions on Biomedical Engineering 53.12 (2006): 2610-2614.

[18] Bin G, Gao X, Yan Z, Hong B. An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method. Journal of neural engineering 6.4 (2009): 046002.

[19] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan T. Brain–computer interfaces for communication and control. Clin Neurophysiol, 113 (6) (2002): 767–791

[20] Chen F, Zhou J, Yang C. Reconstructing orientation field from fingerprint minutiae to improve minutiae-matching accuracy. IEEE Trans. Image Process. 18 (2009): 1665–70

[21] Allison BZ, Neuper C. Could anyone use a BCI?. Brain-computer interfaces. Springer London, 2010. 35-54.

[22] Sellers E, Donchin E. A P300-based brain–computer interface: initial tests by ALS patients. Clinical Neurophysiology, 117 (3) (2006): 538–548

[23] Schettini F., Riccio A., Simione L., Liberati G., Caruso M., Frasca V., Mattia D. Assistive device with conventional, alternative, and brain-computer interface inputs to enhance interaction with the environment for people with amyotrophic lateral sclerosis: a feasibility and usability study. Archives of physical medicine and rehabilitation, 96(3) (2015): S46-S53.

[24] Iturrate I., Antelis J. M., Kubler A., Minguez J. A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation. IEEE Transactions on Robotics, 25(3) (2009): 614-627.