Cognitive Neurodynamics (2023) 17:1401–1416
https://doi.org/10.1007/s11571-022-09901-3

RESEARCH ARTICLE

Design of auditory P300-based brain-computer interfaces with a single auditory channel and no visual support

Yun-Joo Choi[1] · Oh-Sang Kwon[1] · Sung-Phil Kim[1]

Received: 18 April 2022 / Revised: 5 September 2022 / Accepted: 14 October 2022 / Published online: 18 November 2022

Abstract

Non-invasive brain-computer interfaces (BCIs) based on an event-related potential (ERP) component, P300, elicited via the oddball paradigm, have been extensively developed to enable device control and communication. While most P300-based BCIs employ visual stimuli in the oddball paradigm, auditory P300-based BCIs also need to be developed for users with unreliable gaze control or limited visual processing. Specifically, auditory BCIs without additional visual support or multi-channel sound sources can broaden the application areas of BCIs. This study aimed to design optimal stimuli for auditory BCIs among artificial (e.g., beep) and natural (e.g., human voice and animal sounds) sounds in such circumstances. In addition, it aimed to investigate differences between auditory and visual stimulations for online P300-based BCIs. As a result, natural sounds led to both higher online BCI performance and larger differences in ERP amplitudes between the target and non-target compared to artificial sounds. However, no single type of sound offered the best performance for all subjects; rather, each subject indicated different preferences between the human voice and animal sound. In line with previous reports, visual stimuli yielded higher BCI performance (average 77.56%) than auditory counterparts (average 54.67%). In addition, spatiotemporal patterns of the differences in ERP amplitudes between target and non-target were more dynamic with visual stimuli than with auditory stimuli. The results suggest that selecting a natural auditory stimulus optimal for individual users as well as making differences in ERP amplitudes between target and non-target stimuli more dynamic may further improve auditory P300-based BCIs.

Keywords Non-invasive brain-computer interface · Event-related potential · P300 · Auditory brain-computer interface · Sound design

Introduction

Brain-computer interfaces (BCIs) provide alternative means for people to communicate with the external environments without any involvement of motor control, especially for the patients who suffer from neurological disorders such as amyotrophic lateral sclerosis (ALS) and spinal cord injury (Wolpow et al. 2002; Birbaumer and Cohen 2007). Electroencephalography (EEG) has been widely employed as a non-invasive method of sensing brain activities for BCIs, primarily due to its cost-effectiveness and high temporal resolution (Nicolas-Alonso and Gomez-Gil 2012; De Vos 2014). Non-invasive EEG-based BCIs can be categorized into active, passive, and reactive BCIs according to the degree of volitional engagement of BCI users (Zander and Kothe 2011). Among them, reactive BCIs harness neural responses induced by external stimuli to infer users' intentions. Such reactiveness makes BCIs relatively easy to use, even for people who are unfamiliar with the BCI systems, as the users only need to selectively attend to a given stimulus without much mental effort. Event-related potentials (ERPs) and steady-state visually evoked potentials (SSVEPs) are the most prominent EEG patterns exploited by reactive BCIs. In particular, P300 is a key feature for ERP-based BCIs, which is an ERP component elicited approximately 300 ms after a target stimulus onset in the oddball paradigm where an infrequent

✉ Oh-Sang Kwon
oskwon@unist.ac.kr

✉ Sung-Phil Kim
spkim@unist.ac.kr

Yun-Joo Choi
joo618@unist.ac.kr

[1] Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea

target stimulus is randomly given amid a series of frequent stimuli (Donchin et al. 1978; Sara et al. 1994). BCIs based on P300 have been widely investigated and developed into practical systems such as a P300 speller and device control systems (Farwell and Donchin 1988; Carabalona et al. 2010; Corralejo et al. 2014; Kim et al. 2019).

A majority of P300-based BCIs rely on visual stimuli via the oddball paradigm. For instance, the P300 speller, developed by Farwell and Donchin, enables people to select and type a letter by gazing at the flashes in the row and column of the matrix displayed on the screen (Farwell and Donchin 1988). However, P300-based BCIs based on visual stimuli can be limited in certain circumstances. First, some users who have a deficit of eye movements due to severe disabilities may find it difficult to control their gazes at a target. For example, patients with late-stage ALS may develop completely locked-in syndrome (CLIS) and have unreliable gaze control (Harvey et al. 1979; Hayashi and Kato 1989; Kübler and Birbaumer 2008). Second, the integration of BCIs with augmented reality (AR) and virtual reality (VR) technologies can bring limitations to visual P300-based BCIs; while the development of P300-based BCIs in AR and VR presents visual stimuli in various environments instead of those limited to traditional monitors, it can also cause a visual distraction that might induce irrelevant ERP components and consequently lower the BCI performance (Takano et al. 2011; Zeng et al. 2017; Si-Mohammed et al. 2020).

To overcome these limitations of visual P300-based BCIs, other approaches to designing P300-based BCIs without visual stimuli have been investigated, including auditory P300-based BCIs (Furdea et al. 2009; Klobassa et al. 2009; Simon et al. 2015; Oralhan, 2019). A key aspect that differentiates auditory P300-based BCIs from their visual counterparts, is the design of stimuli, as the oddball paradigm per se is supramodal (Huang et al. 2018). A variety of auditory stimulation design methods have been proposed for P300-based BCIs: some proposed using both auditory and visual stimuli together by employing supportive visual cues in addition to auditory stimuli. For example, in the study of Furdea et al. (2009), a visual matrix that did not flash was provided additionally as the reference in the design of a P300 speller with auditory stimuli. Klobassa et al. (2009) also used a static character matrix with auditory stimuli for a P300 speller where the rows and columns of the visual matrix were assigned to different environmental sounds. While visual matrices did not provide any stimulation in these cases, they effectively supported the users to remember which auditory stimuli they needed to attend to (Furdea et al. 2009). Moreover, other methods used both supportive visual matrices and multiple auditory channels. For example, Simon et al. (2015) used five speakers to present different animal sounds along with a static visual support matrix, in which each row and column was coded with the animal tone. Oralhan (2019) also introduced an auditory P300 speller, which presented spatially localized auditory stimuli via two speakers with corresponding visual references on the screen. While these designs of auditory stimulation mixed with visual references exhibit a clear purpose of improving the relatively low performance of auditory P300 BCIs by helping the users process the auditory stimuli more effectively through visual references, a new design of auditory stimulation is necessary for certain real-life environments with no visual support available (e.g., using BCIs during driving a car where gazing at distracting visual stimuli is dangerous, or using BCIs to control smart home systems during chores where visual processing is concentrated on given tasks). In addition, it will provide greater flexibility to the design of BCIs if multiple auditory stimuli can be delivered through a single channel, as multiple sound sources can be limited in cost and portability.

For such auditory P300-based BCIs without additional visual cues or sound localization, the selection of auditory stimuli becomes crucial. To achieve high accuracy and fast information transfer rate (ITR) with a serial presentation of auditory stimuli through a single channel, it is necessary to present stimuli that can be easily distinguished from one another as well as transmitted rapidly in a short time. The speed-accuracy tradeoff should be taken into special account in this case, because auditory stimuli with a longer duration may be more accurately distinguished while reducing ITR at the same time. Previous studies of auditory BCIs have investigated various types of auditory stimuli. The studies by Furdea et al. (2009) and Oralhan (2019) used acoustically presented numbers via the human voice. Also, Klobassa et al. (2009) presented environmental sounds while Halder et al. (2016) used beeps. Meanwhile, Höhne et al. (2012) demonstrated that natural sounds would be more suitable than artificial tones for auditory BCIs by showing that the stimuli of spoken and sung syllables not only showed high ergonomic ratings but also improved performance compared to artificially generated tones. In addition, Huang et al. (2018) indicated that both the accuracy and ITR of online BCIs were higher with natural drip-drop sounds compared to the ones with beeps. While these studies indicated that auditory BCIs could benefit from natural sounds than artificial sounds, little is known about which type of natural sound is more appropriate for auditory BCIs.

It has been consistently reported that auditory P300-based BCIs yielded poorer performance than their visual counterparts (Furdea et al. 2009; Belitski et al. 2011; Oralhan 2019). Studies showed higher accuracy using the visual P300 spellers than using the auditory ones [e.g., 94.62% vs. 65.00% (Furdea et al. 2009), or 78.06% vs.

54.08% (Oralhan 2019)), as well as higher ITR (e.g., 6.80 bits/min vs. 1.54 bits/min (Furdea et al. 2009), or 5.17 bits/min vs. 3.43 bits/min (Oralhan 2019)]. The higher performance of visual spellers reportedly accompanied higher P300 peak amplitudes (Oralhan 2019; Belitski et al. 2011). However, it remains unknown whether P300 amplitudes would differ among different natural and artificial sounds. As the higher performance of visual BCIs entailed higher P300 amplitudes, higher P300 amplitudes are expected with natural sounds if they lead to better performance than artificial sounds. Identifying differences in ERP waveforms between auditory and visual stimulations will be important to understand why visual P300-based BCIs outperform auditory ones and potentially how to reduce the performance gap.

This study aimed to investigate the optimal design of auditory stimuli for P300-based BCIs by comparing different types of natural sounds as well as artificial ones regarding online BCI performance. Auditory stimuli were presented serially through a single auditory channel with no visual support. Individual differences in the design of optimal auditory stimuli were also investigated. Then, we compared the online performance of auditory P300-based BCIs with individually selected auditory stimuli to that of visual P300-based BCIs. Moreover, differences in the ERP waveforms elicited by each stimulation were examined. To compare stimulations for P300-based BCIs in a real-life environment, we used BCIs built to control the functions of a home appliance—an electric light (EL) device—in real-time.

Among many natural sounds that have been used for auditory P300-based BCIs—including the human voice, animal sounds, environmental sounds such as a bell, bass, ring, thud, chord, buzz, drip drops, and others (Klobassa et al. 2009; Belitski et al. 2011; Simon et al. 2015; Huang et al. 2018)—we chose human voice and animal sounds for this study. Previous studies showed that spoken words have been claimed to provide straightforward stimuli and require less training time needed to run auditory BCIs (Ferracuti et al. 2013). Other studies on auditory BCIs also showed relatively high performance with human voice stimuli (Furdea et al. 2009; Belitski et al. 2011; Höhne et al. 2012; Chang et al. 2013; Oralhan 2019). In addition, animal sounds render higher discriminability than artificial tones (Simon et al. 2015), according to follow-up studies (Baykara et al. 2016). For the comparison with natural sounds, we selected a beep as an artificial sound, as it has been often used for auditory P300-based BCIs (Halder et al. 2010; Höhne et al. 2011; Huang et al. 2018).

Methods

Participants

Thirty healthy subjects (14 females aged 19 to 37 with a mean of 23.6 ± 3.83) participated in the main study, while a separate group of six healthy subjects (4 females aged 21 to 26 with a mean of 23.5 ± 1.87) participated in the preliminary behavioral study. No subject reported suffering from any neurological or psychiatric disorders or hearing impairment, and all of them had a normal or corrected-to-normal vision. All subjects gave informed consent for the study, approved by the Ulsan National Institute of Science and Technology, Institutional Review Board (UNIST-IRB-21-22-A).

Auditory stimuli

Three types of auditory stimuli were designed for the experiment, including the human voice, animal, and beep sounds. Furthermore, the voice and animal sounds were categorized as natural sounds, while the beep was categorized as an artificial sound. In each stimulus type, we designed four stimuli that were supposed to be distinguished from one another, except for the animal sound type, where we initially designed six sounds and later selected four via a preliminary behavioral study (see below). This additional selection procedure for animal sounds was needed, as animal sounds could not be created with the adjustment of sound parameters unlike the human voices and beeps, thus opting to select animal sounds empirically. The number of stimuli (4) was determined based on the number of functions of an electronic light (EL) device (Philips hue 2.0, Philips, Netherlands) that were controlled by a P300-based BCI system in this study. Every sound stimulus used in this study was equalized to have identical loudness by adjusting the root mean square (RMS). Also, the duration of every stimulus was set as 275 ms. Identical stimuli were presented via earphones (AZLA, Republic of Korea) plugged into both ears. The spectrograms of every stimulus are depicted in Fig. 1.

For the type of beep sounds, four beeps with different frequencies were designed. According to the previous study that investigated auditory P300-based BCIs using beeps (Huang et al. 2018), we created three beep sounds with frequencies of 800, 1000, and 1200 Hz, respectively. We also added another beep with a frequency of 1400 Hz to match the number of stimuli required for our BCI experiment.

For the type of voice sounds, we created four stimuli by recording human voices reading Korean words, ‘han’, ‘dul’, ‘set’, and ‘net’, which mean the numbers 1, 2, 3, and

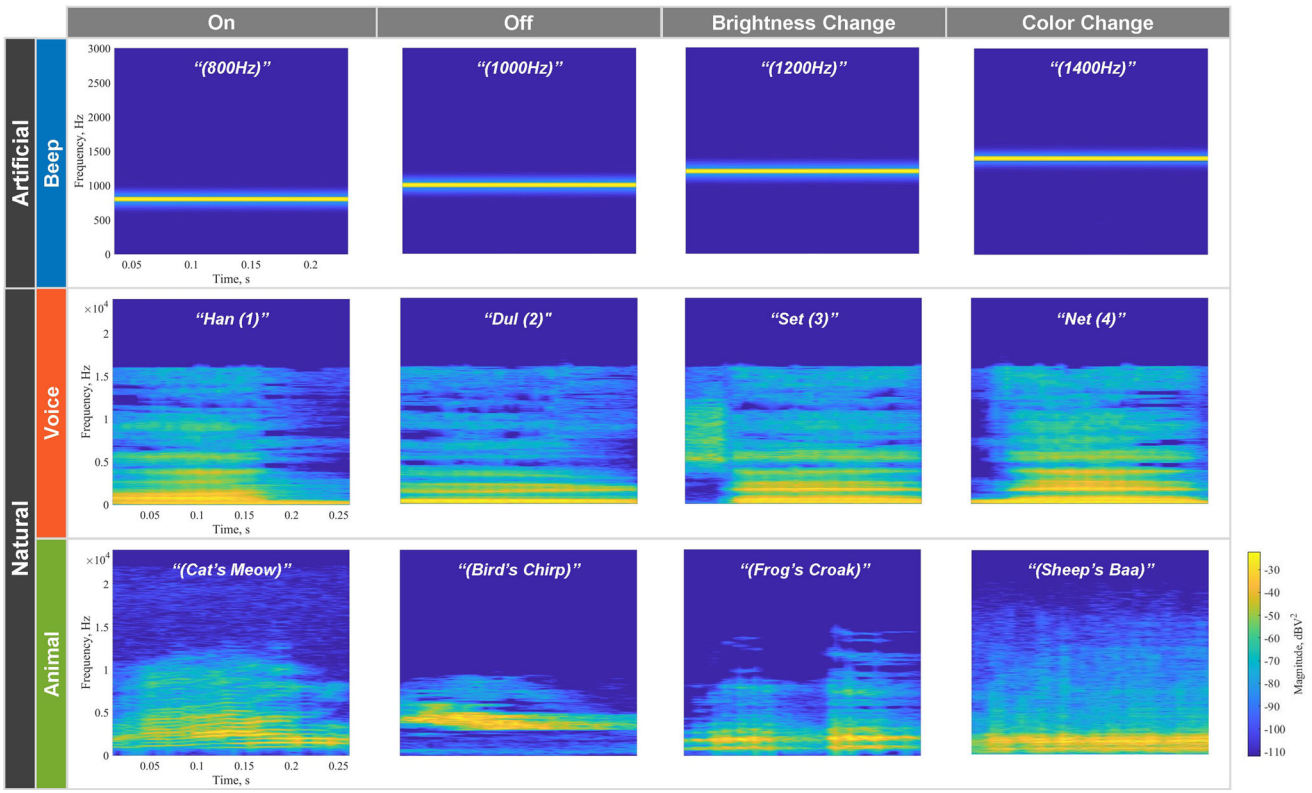

Fig. 1 Spectrograms of each auditory stimulus

4, respectively. 'Hana' is the original Korean word meaning the number 1, but it was reduced to 'han' to equalize the syllable length to one. We intended to create a single-syllable sound to accommodate a short stimulation period. The human voice reading each function of the electric light device could be another possible design, but it was difficult to find an appropriate single-syllable word to effectively represent each function (e.g. 'brightness change', 'color change'). Additionally, some verbal sounds were not easily distinguishable when they were presented shortly (e.g. 'on' for 'light on', 'off' for 'light off'). A male voice with the same pitch was recorded and edited to make the stimuli.

For the type of animal sounds, a cat's meow, bird's chirp, frog's croak, dog's bark, duck's quack, and sheep's baa were initially chosen. The original sound clips were downloaded from the websites, https://mixkit.co/ and https://www.epidemicsound.com/, and edited to make the length and RMS of the sounds equal. We posited that if one sound is more salient than others, it would attract more attention regardless of its position as a target or non-target and therefore not be a desirable stimulus. To select and remove those salient sounds among the six sounds, we conducted a preliminary behavioral study on six subjects. The behavioral study consisted of three sessions. In each session, every possible pair of the six sounds were presented randomly to the subjects. Two sounds of each pair were given successively, and the subjects were asked to press 1 if the first clip sounded more salient and 2 if vice versa. If the two clips sounded with a similar level of saliency, the subjects pressed 3. This session was repeated three times, and a total of 90 pairs of the sounds were evaluated. As a result, the number of responses that selected dog's bark and duck's quack as the salient clips was the highest among six clips on average, which was also consistent across sessions (Online Resource 1). Consequently, we discarded these two clips and selected the rest for the experiment: cat's meow, bird's chirp, frog's croak, and sheep's baa.

Experimental protocol

For each subject, the experiment was conducted on two different days with an interval of 3 to 7 days (mean 5.23 ± 1.63 days) in between. On the first day, three different auditory P300-based BCI systems with each stimulus type were operated in a randomized order. A post-survey was conducted after each auditory BCI system. On the second day, a visual BCI system, as well as an auditory BCI system with the stimulus type that had shown the best BCI control accuracy on the first day were operated in a randomized order (Online Resource 1). On both days, paper-based instructions were shared at the beginning of

the BCI experiment with each stimulus type, which described the association of each auditory stimulus with the function of the device. After the subjects read the instructions, they performed a pre-training task (see below) before the online operation of P300-based BCIs. The pre-training task was obligatory only on the first day but optional on the second day.

Pre-training task

Before the online BCI operation, the pre-training task was prepared to allow the subjects to become familiar with the auditory stimuli. This pre-training was implemented according to the previous study's report on the effects of familiarity on the performance of auditory BCIs (Baykara et al. 2016). It demonstrated that familiarity with auditory oddball paradigms should be considered, as high task demands are needed in an oddball task with a series of rapidly presented auditory stimuli. Since the subjects in the present study operated multiple BCIs with different types of stimuli sequentially, familiarity could confound BCI performance. Thus, we prepared pre-training to alleviate the effect of familiarity on BCI performance.

For each type of auditory stimuli, the pre-training task started by presenting each of the four stimuli successively via earphones (Online Resource 1). Then, the pre-training task blocks followed. In each block, a target stimulus was presented at first via the earphones, followed by the task instruction displayed on the monitor (1920 × 1080 resolution, Full-high-definition (FHD), LG Electronics Co., Ltd., Republic of Korea), which asked the subjects to count covertly the number of times the target sound was heard in the subsequent presentation of a series of stimuli. Then, each of the four stimuli was presented randomly for 275 ms for 6 to 8 repetitions with a 250 ms inter-stimulus interval. A fixation (white cross) was continuously shown on the monitor during the presentation of all stimuli. After the presentation of stimuli, the subjects entered their counting result on the computer keyboard. Visual feedback was provided on the monitor about whether the subject's count was correct or not. Afterward, the subjects were asked whether they wanted to continue with the next block or finish the pre-training task. As the pre-training task was designed only to let the subjects become familiar with the sounds used in the experiment, the subjects could finish the task whenever they felt familiar with the stimuli. Although we did not conduct any explicit test on each subject's familiarity, we provided instruction about two basic criteria to the subjects: first, the subjects should conduct a minimum of five blocks; and second, they should give a correct answer in at least one block. Accordingly, the number of blocks differed across individual subjects as well as stimulus types. Overall, eight blocks of the pre-training task were conducted on average.

On the second day, pre-training was optional, as we assumed that the subjects were already familiar with the stimuli. If the subjects opted for pretraining, the two criteria above were not applied, and the subjects could conduct the pre-training blocks as many as they wanted. Only eight of the thirty subjects chose to perform the pre-training task on the second day.

Online BCI operation

The main task was to control an EL device using a P300-based BCI system. EL was located in front of the subjects so that they could receive the closed-loop feedback about the brain control of EL. We built three BCI systems according to the stimulus type. In each system, each of the four stimuli was associated with four different functions of EL: 'light on', 'light off', 'brightness change', and 'color change' (Fig. 2). In the auditory BCI system, the stimuli were presented via earphones, and an additional USB sound card (COMSOME SD-30 T) was used to reduce jitter. A white fixation cross was presented on the monitor (1920 × 1080 resolution, Full-high-definition (FHD), LG Electronics Co., Ltd., Republic of Korea) to minimize the eye movement of the subjects, without any other visual stimuli. In the auditory BCI systems, a target function of EL that the subjects needed to select was provided by the auditory stimulus sound associated with that target function.

In contrast, in the visual BCI system, the four visual stimuli were simultaneously displayed on the monitor. Each stimulus was designed as a blue square containing an icon that directly described an associated function of EL. The stimuli were placed in every corner of the screen, and the length of the square was a fourth of the vertical length of the screen. To instruct the target, the border of the corresponding square turned to red. Each of the four stimuli was presented one at a time by changing its color from blue to green.

There were 30 blocks in the training session and 15 blocks in the testing session for every auditory or visual BCI system. A block contained 28 trials of the presentation of a stimulus in a random order where each of the four stimuli was presented 7 times. An auditory or visual stimulus was presented for 275 ms with an ISI of 250 ms, making the duration of a trial 525 ms. Thus, a block spanned 14,700 ms and an inter-block interval was approximately 5000 ms. A target stimulus was informed to the subjects before each block began (see above). The subjects were instructed to count covertly the number of target presentations in both training and testing sessions. Since the primary goal of counting was to maintain the

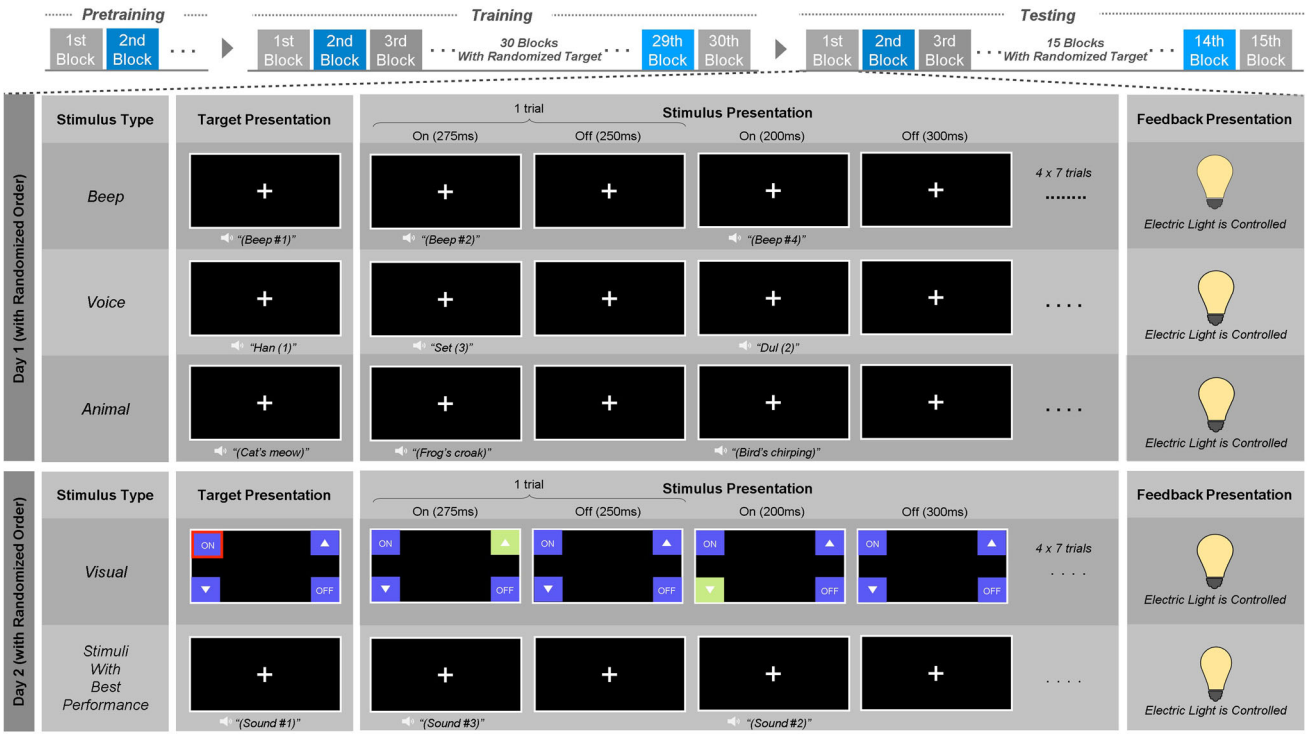

Fig. 2 Experimental protocol

subjects' attention to a target stimulus, we did not verify the correctness of counting at the end of every block to save experiment time. At the end of a block in the test session, EL was controlled by the control output from the BCI system, whereas no explicit feedback was given to the subjects during the training session.

Post-survey

On the first day, a post-survey was conducted after the online operation of each auditory BCI system. As the survey asked about the experience of the auditory BCI, it was not conducted on the second day. The survey consisted of six questions. The first four questions were based on the NASA Task Load Index (NASA-TLX, NASA Human Performance Research Group, 1987), regarding the subjective workload of mental demand, performance, effort, and frustration. Each question was evaluated on a 7-point scale ranging from 'very low' to 'very high.' Lastly, after all the post-surveys were conducted, the subjects were asked to rank three auditory stimuli according to their suitability for real-life BCI systems (see Online Resource 1 for the list of the questionnaires used in the post-survey).

Data acquisition and preprocessing

The scalp EEG signals were acquired using 31 active wet electrodes (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, CP5, T8, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, and O2) with a standard EEG cap following the 10–20 system of American Clinical Neurophysiology Society Guideline 2. The electrode attached to the mastoid of the left ear was used as the ground, while that of the right ear as the reference. The impedance of all electrodes was kept under 5 kΩ. EEG signals were amplified and sampled at 500 Hz using a commercial EEG amplifier (actiCHamp, Brain Products GmBH, Germany).

The sampled raw EEG signals were processed through the following procedure: (1) High-pass filtering above 0.5 Hz was applied; (2) For each channel, if more than 70% of all other channels exhibited a cross-correlation lower than 0.4 with that channel after band-pass filtering between 0.5 and 1 Hz, the channel was deemed as a bad one and was removed (Bigdely-Shamlo et al. 2015); (3) Potential noise components from the reference were removed using the common average reference (CAR) technique; (4) The re-referenced signals were low-pass filtered below 50 Hz; (5) Artifacts were minimized using the artifact subspace reconstruction (ASR) method (Bigdely-Shamlo et al. 2015; Chang et al. 2020; Mullen et al. 2015); and (6) Low-pass filtering below 12 Hz was applied to the signals for the ERP analysis. The average number of bad channels was two across all the BCI systems.

Feature extraction

We extracted EEG features for classification to discriminate a target from non-target stimuli. Features were composed of the ERP amplitudes from every channel. To obtain ERPs induced by each stimulus, we adaptively determined the epoch of ERPs for each subject, as the previous study showed that there are individual variations in response time to auditory stimuli (Ng and Chan 2012). Moreover, the time taken for information processing would be different between visual and auditory stimuli because auditory stimulation delivered the cue information continuously via sound waves, whereas visual stimulation simply altered the color at stimulus onset, as supported by the discrepancies reported in previous studies regarding response time between visual and auditory stimulations (Cheng et al. 2008; Ng and Chan 2012).

In this study, we determined the epoch of ERPs based on the accuracy of classifying between target and non-target in the training session. With the training data, eight options of the post-stimulus epoch length along with a fixed-length baseline were compared using cross-validation, where eight options included 800, 900, 1,000, 1,100, 1,200, 1,300, 1,400, and 1500 ms. and the baseline was defined as − 200 to 0 ms to stimulus onset. As the P300 component would be elicited 300–600 ms after stimulus onset, we set the shortest option as 800 ms to sufficiently embrace the P300 component (Puanhyuan and Wongsawat 2012). ERPs were obtained by averaging EEG amplitudes within a given epoch length across trials and baseline-corrected by subtracting the averaged amplitude of the baseline from each data point. With these ERP amplitudes as features, we classified target versus non-target using a support vector machine (SVM) classifier (see below for details) with a leave-one-block-out validation scheme in which 29 blocks were used for training and 1 block for validation repeatedly. The epoch length option with the highest validation accuracy was chosen as the final epoch length for each subject.

Afterward, a feature vector was constructed as follows: (1) A set of amplitude values from 200 ms after stimulus onset to an epoch length was extracted for each channel. We excluded a period from 0 to 200 ms after stimulus onset as we focused on the endogenous ERP components that occur relatively slowly after stimulus onset (e.g. N2pc, P300). The number of features per channel (N_C) was $N_C = L \times f_s$, where L is the epoch length in second and f_s is the sampling rate (500 Hz); (2) N_C features from each channel were then concatenated to create a feature vector of the dimensionality $N_f = N_C \times N_e$, where N_e is the number of channels excluding bad channels. The average N_f across the subjects was 422 (average number of features per channel with varying epoch length) × 29 (average number of channels after bad channel removal) = 12,238. (3) Down-sampling by a 10-sample window with a factor of 2 was applied to extracted features to reduce dimensionality, resulting in N_f = 84.4 (average number of features per channel) × 29 (average number of channels) = 2447.6.

Classification and evaluation

We built a binary classifier based on a linear kernel SVM to discriminate a target from non-target stimuli. This type of classifier was chosen following our previous studies of the P300-based BCIs to control home appliances (Kim et al. 2019). As a single block produced four ERP feature vectors in response to four stimuli, one target, and three non-target, an input feature matrix, $D \in R^{N_d \times N_f}$ was constructed for SVM training, where N_d is the number of data samples. From 30 blocks of the training session, N_d = 120 training samples were obtained, including 30 samples for target and 90 for nontarget, respectively. In each block of the test session, four ERP feature vectors were collected and classified as either target or non-target. Specifically, let X_C be a feature vector corresponding to the c-th stimulus in a block. The SVM classified X_C with a penalty parameter fixed as 1 and produced the classification score, $f(X_C)$, which represents the probability that the c-th stimulus was a target. Finally, the target stimulus, T was determined by: $T = argmax_C f(X_C)$. Then, the function associated with the target stimulus was executed to control EL. When training the classifier, the number of features exceeded the number of training samples; however, as our previous study successfully classified such a large number of ERP features using SVM for online BCI control (Kim et al. 2019), we followed a similar feature extraction procedure without further feature selection steps.

The performance of online BCI systems was assessed by target detection accuracy = $\frac{N_C}{N_T}$, where N_C is the number of correctly selected testing blocks and N_T is the total number of testing blocks. As there were 15 blocks in the test session for each BCI system, N_T was 15 in this study. In addition, ITR was calculated to evaluate online BCI performance:

$$\mathrm{ITR} = \frac{N + PP + (1-P)\left(\frac{1-P}{N-1}\right)}{T}. \quad (1)$$

To compare the classification accuracy and ITR among BCI systems, statistical tests of rmANOVA followed by a post-hoc paired t-test with Bonferroni correction were conducted.

Post-hoc ERP analysis

We compared ERP waveforms induced by target and non-target stimuli among different BCI systems. We assessed ERP waveforms in response to each stimulus type using the training data. For the comparison of ERP waveforms, the epoch length was equalized as 800 ms for all subjects, as cross-validation outcomes were not substantially different across various epoch lengths, thus we opted to use the shortest length for the simplicity of analysis (Online Resource 1). In ERP waveforms, we first analyzed the P300 component determined as a positive component appearing between 150 and 600 ms after stimulus onset. We identified the peak of the P300 component at each channel in each subject as the highest amplitude within this designated time window and the latency as the time when this highest amplitude occurred.

We also analyzed a difference in the amplitudes of ERP waveforms between target and nontarget stimuli. We assumed that a bigger difference between target and non-target would be likely to make feature vectors more distinguishable, thus potentially illuminating the effect of the stimulus type on BCI performance. In addition, we conducted two sample t-tests with multiple comparison correction by False Discovery Rate (FDR) to identify time points at each channel that show significant differences in ERP amplitudes between the target and non-target, which resulted in t-value maps over time and channel for each BCI system. Note that the channels in which a bad channel was detected at least in one subject were removed from this post-hoc analysis. As a result, as FT10 was removed in most subjects, FT10 was completely excluded from the ERP analysis.

Results

Comparison of online BCI performance between different auditory stimuli

The average (± standard deviation) classification accuracy from the online BCI operation on the first day with the beep, voice, and animal sounds were 31.56% (± 13.55), 49.11% (± 22.11), and 55.78% (± 21.12), respectively (Fig. 3a). Note that the chance level was 25% as there were four possible selections by the BCI systems. The rmANOVA test revealed a significant difference in accuracy among three auditory BCI systems (F (2,58) = 21.79, $p < 0.001$). A post-hoc t-test with Bonferroni correction showed significant differences between the beep and voice or between the beep and animal sounds ($ps < 0.001$).

When operating each BCI system, all the subjects (N = 30) showed better performance with natural sounds than with artificial sounds. Specifically, 20 subjects showed the highest accuracy with the animal sounds while 10 subjects showed the highest accuracy with the voice sounds. In terms of the worst performance, 6 subjects showed the lowest accuracy with the voice, 1 with the animal sounds, and 23 with the beeps. Notably, when the number of repetitions of the stimulations increased, accuracy also increased with the voice and animal sounds, but not with the beeps (Fig. 3b, Table 1).

The average values (± standard deviation) of ITR from the online BCI operation with the beep, voice, and animal sounds were 0.32 bits/min (± 0.36), 1.43 bits/min (± 1.91), and 1.83 bits/min (± 1.88), respectively (Fig. 3c). The rmANOVA test on ITR showed a significant difference among three auditory BCI systems (F (2,58) = 8.88, $p < 0.001$). A post-hoc t-test with Bonferroni correction showed a significant difference between

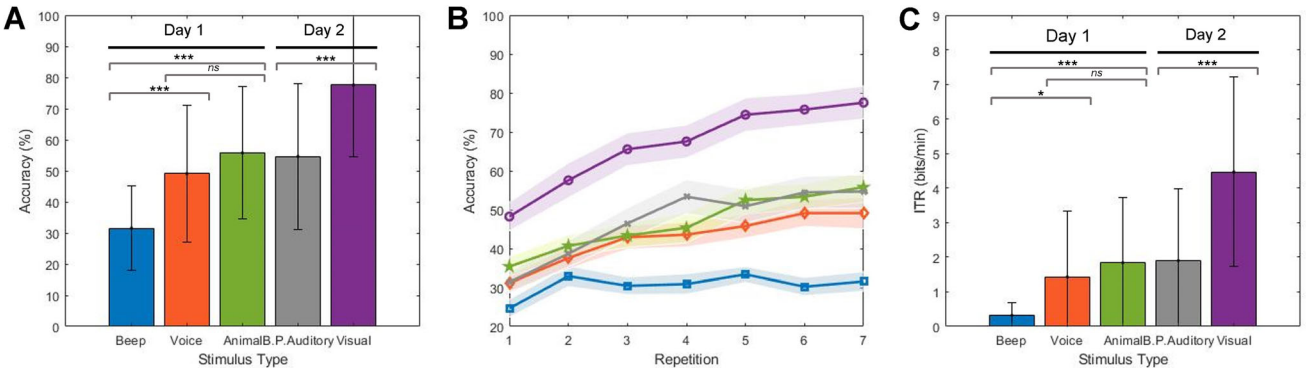

Fig. 3 Performance shown in online BCI session. **a** Averaged classification accuracy (%) and standard error of mean depicted with bar graphs and the vertical lines, respectively. **b** Averaged values of the accuracy (%) per each type of the stimuli when the repetition number of the stimulation was differed. **c** Averaged values of ITR (bits/min) per each type of the stimulus (*$p < 0.05$; **$p < 0.005$; ***$p < 0.001$)

Table 1 Averaged classification accuracy (%) per each group of the stimulus when the repetition number of the stimulations was differed

Rep no	1	2	3	4	5	6	7
Beep							
Mean	24.67	32.89	30.44	30.89	33.33	30.22	31.56
STD	12.64	14.43	12.22	14.17	10.93	12.59	13.55
Voice							
Mean	31.11	37.56	42.89	43.56	45.78	49.11	49.11
STD	11.39	15.21	16.76	17.31	16.68	18.65	22.11
Animal							
Mean	35.33	40.67	43.33	45.33	52.44	53.33	55.78
STD	14.13	15.20	16.77	21.26	19.24	17.77	21.12
B.P. auditory							
Mean	31.33	38.67	46.44	53.33	50.89	54.44	54.67
STD	15.18	18.81	21.76	23.88	23.71	22.63	23.45
Visual							
Mean	48.22	57.56	65.56	67.56	74.44	75.78	77.56
STD	20.17	24.48	23.10	23.21	23.56	21.83	23.06

unit: %

the beep and voice or between the beep and animal ($ps < 0.001$).

Individual preference of auditory stimuli

Based on the online BCI performance results above showing that BCIs with either of the two natural sounds yielded higher accuracy than those with the artificial sound and that all the subjects achieved the best BCI performance with natural sounds, we further investigated if natural sounds provided more suitable stimuli for BCIs than artificial ones in general. If that would be the case, we posited that a performance margin between the two natural sounds would be smaller than that between the natural sound with lower performance and the artificial sound. In each subject, we first calculated a difference in accuracy between the two natural sounds (dNN) and then a difference in accuracy between one of the natural sounds yielding lower performance and the beep sound (dNA) (Fig. 4a). The average values of dNN and dNA were 23.56% (± 12.47) and 9.11% (± 12.68), respectively (Fig. 4b). Only 7 out of 30 subjects showed larger values of dNA than dNN. Paired t-test showed that dNN was significantly larger than dNA (t (29) = 3.74, $p < 0.001$). This result was in contrast to our expectation that natural sounds would be generally more suitable for auditory P300-based BCIs than artificial sounds. Rather, it indicated that the selection of a specific type of natural sounds preferred by each user may be more important for auditory P300-based BCIs.

Comparison of ERPs among auditory stimuli

We analyzed ERPs induced by the target and non-target of each auditory stimulus type (see Fig. 5a for the grand average ERPs at representative channels). The peak latency and amplitude of P300 components for the target among three auditory stimulus types were compared at each channel using rmANOVA (Online Resource 1). In comparison to the peak amplitude, it was revealed that significant differences were observed at 10 channels (Fz, T7, C4, CP5, CP2, P7, Oz: $p < 0.05$, FC2, Cz, O1: $p < 0.001$). A post-hoc t-test with Bonferroni correction on these channels further revealed that the peak amplitude induced by the beeps was smaller than that by the voices at 6 channels (Fz, FC2, Cz, C4, CP2, O1), and that by the animal sounds at 7 channels (FC2, T7, Cz, CP5, P7, O1, Oz). Meanwhile, only one channel (T7) showed that the peak amplitude induced by the voices was smaller than that by the animal sounds ($p < 0.05$). In comparison to the peak latency, there were only four channels (C4, P8, Oz, O2) that showed significant differences among three auditory stimulus types ($ps < 0.05$) and only two channels (T7, O2) showed significant differences in the post-hoc t-tests. These channels showed that the peak latency of the voice sounds was shorter than that of the animal sounds ($ps < 0.05$). Table 2 summarizes the significant results with the corresponding p values.

As larger differences in the ERP amplitudes between the target and non-target would be linked to BCI performance, we examined those differences among the auditory stimulus types. To this end, we analyzed the ERP amplitudes for

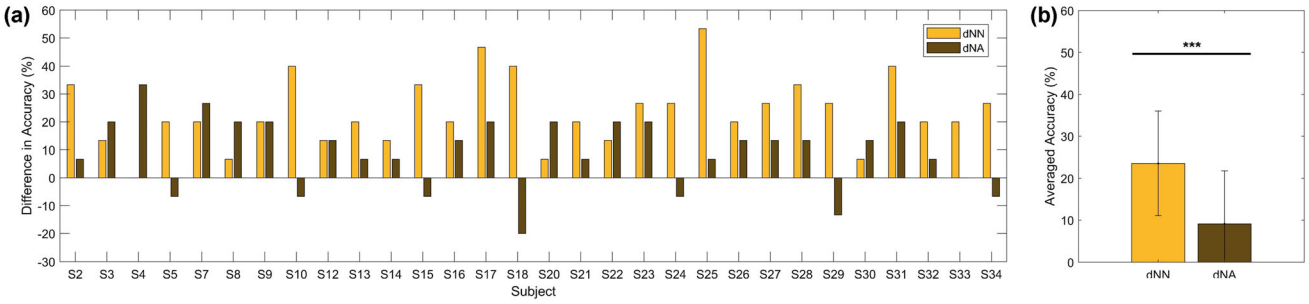

Fig. 4 **a** Individual values of dNN and dNA, **b** averaged values of dNN and dNA (***$p < 0.001$)

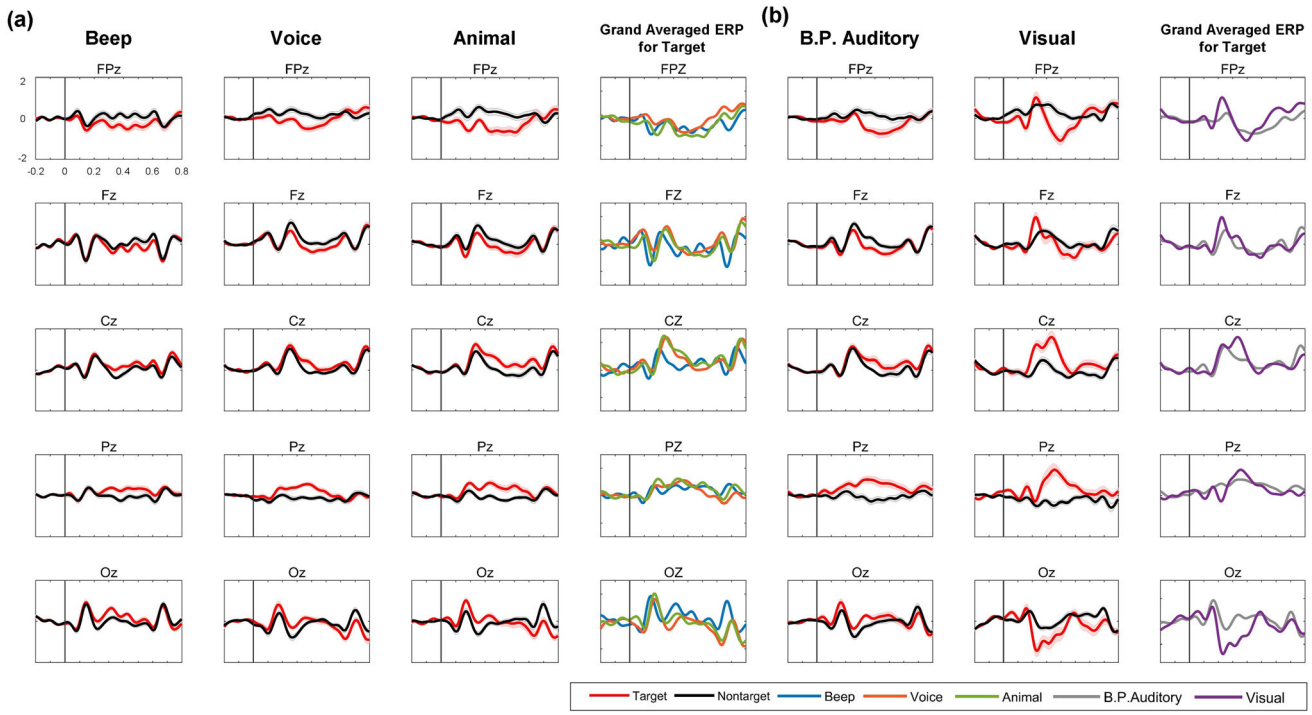

Fig. 5 Grand averaged ERP waveforms across all subjects shown in five channels (FPz, Fz, Cz, Pz and Oz). **a** ERP waveforms shown with beep, human voice, and animal sounds. **b** ERP waveforms shown with the auditory and visual stimulations on the second day. Their last figures were depicted to compare the ERPs of the target

Table 2 p values resulted by the post-hoc Bonferroni tests, which were conducted to compare peak amplitude and latency of P300 components only with the channels that showed significant differences in rmANOVA (light-brown: $*p < 0.05$, pink: $**p < 0.005$, dark-brown: $***p < 0.001$)

	Peak amplitude										Peak latency			
Ch	Fz	FC2	T7	Cz	C4	CP5	CP2	P7	O1	Oz	F4	T7	Oz	O2
Beep < voice	0.03	0.000	1.000	0.006	0.02	1.00	0.02	0.44	0.01	1.00	0.05	1.00	1.00	1.00
Beep < animal	0.54	0.002	0.017	0.003	0.17	0.03	0.10	0.03	0.000	0.005	0.43	0.12	0.08	0.07
Voice < animal	0.78	1.00	0.012	0.812	1.00	0.10	1.00	0.44	1.00	0.44	0.41	0.02	0.13	0.03

target and non-target in each subject, which were obtained by averaging over training blocks. Additional averaging of the ERP amplitudes for non-target was conducted over three non-targets. First, we calculated the differences between these two ERP amplitude waveforms for each stimulus type in each subject (see Fig. 6a for the grand averaged ERP differences at representative channels) and constructed topographies of them over time (Fig. 6b). Overall, relatively larger differences were observed with the natural sounds compared to the beeps. Also, negative difference values were largely observed in frontal areas whereas positive values were observed in parietal and occipital areas (Fig. 6a, b). We compared the area between 200 and 800 ms of the ERP difference waveforms among three stimulus types using rmANOVA and observed significant differences at 23 channels (see Table 3). A post-hoc t-test with Bonferroni correction showed that the area of the beeps was smaller than those of the human voice or animal sounds at 19 channels, while there was no significant difference between the two natural sounds at any channel (Table 3). The topographies also displayed larger differences between target and non-target with the human voice and animal sounds than with the beeps, especially in the centro-parietal areas at 400 ms after stimulus onset (Fig. 6b). Moreover, such larger differences were continuously present 500–600 ms after stimulus onset with animal sounds.

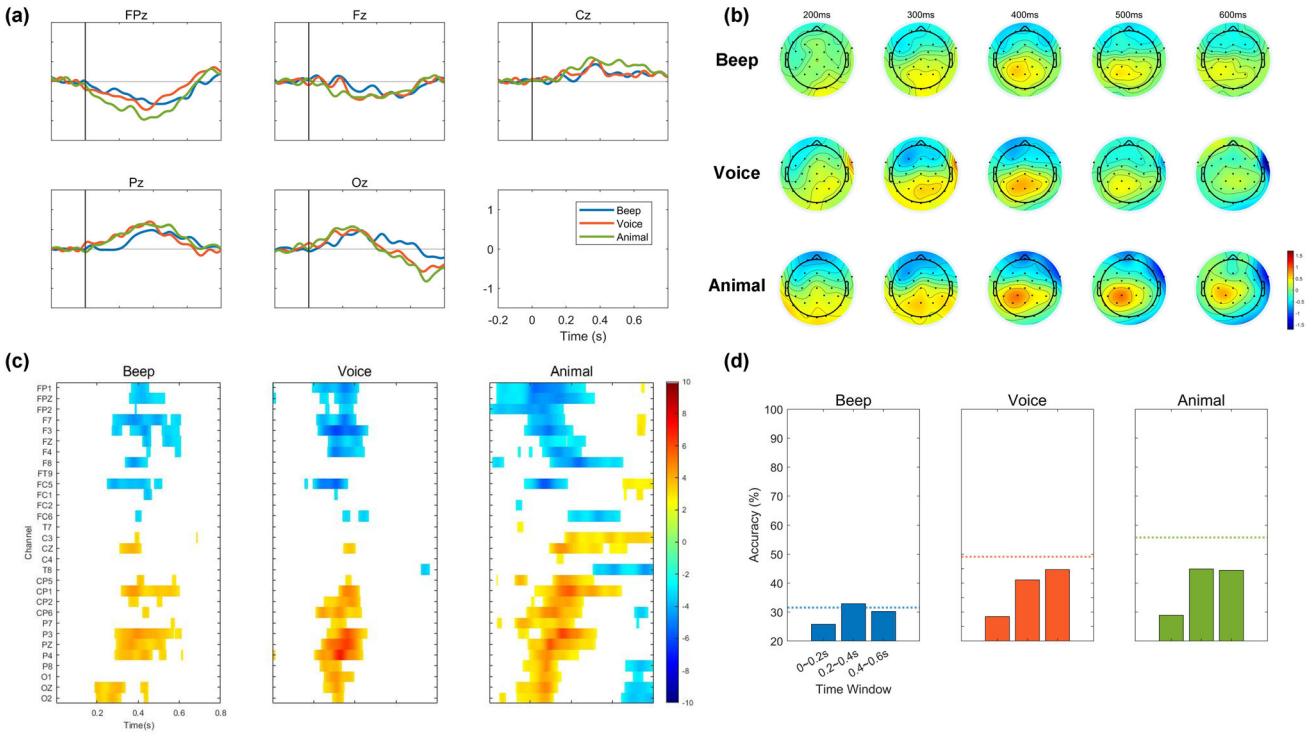

Fig. 6 **a** Grand averaged waveforms and **b** topographies showing the differences in ERP amplitudes of the target and nontarget induced by three different auditory BCI systems. **c** Time-channel maps with t-values of each feature determined by two-sample t-tests, which were conducted to identify the features that showed significant differences between the target and nontarget. The features with positive t-values represent that the ERPs induced by the nontarget were bigger than the ones of the targets and they are colored in yellow or red. **d** Averaged classification accuracy derived by using the data of each different time window with the horizontal dotted lines representing the mean accuracy of the online tests

Table 3 p values and F-values resulted by rmANOVA, which was conducted to compare the area of the ERP waveforms showing the differences between the ERP amplitudes of the target and nontarget between 200 and 800 ms, on the first line and the ones resulted by the post-hoc. Bonferroni tests on the other three lines (light-brown: $*p < 0.05$, pink: $**p < 0.005$, dark-brown: $***p < 0.001$)

Ch		FP1	FPz	F3	Fz	F4	F8	FC1	FC2	FC6	T7	C3	Cz	T8	CP5	CP1	CP2	CP6	P7	P3	Pz	P8	O1	Oz
rmANOVA	p-value	0.04	0.05	0.001	0.000	0.004	0.03	0.000	0.000	0.01	0.000	0.000	0.000	0.000	0.001	0.002	0.01	0.01	0.004	0.02	0.01	0.01	0.004	0.03
	F-value	3.31	3.19	7.56	13.65	5.95	3.82	15.06	27.43	4.58	10.46	10.17	14.61	13.47	7.41	6.72	5.52	5.72	6.14	4.28	5.09	4.56	5.97	3.76
Post-Hoc (p-value)	Beep – voice	0.19	0.34	0.01	0.001	0.04	0.09	0.000	0.000	0.04	0.003	0.01	0.001	0.001	0.03	0.22	0.05	0.12	0.02	0.48	0.32	0.11	0.05	0.17
	Beep – animal	0.12	0.09	0.02	0.000	0.02	0.01	0.000	0.000	0.04	0.001	0.003	0.000	0.001	0.01	0.004	0.03	0.03	0.03	0.04	0.01	0.05	0.04	0.10
	Voice – animal	1.00	1.00	0.24	1.00	0.98	1.00	0.83	0.57	0.78	1.00	1.00	1.00	0.42	1.00	0.20	1.00	0.64	1.00	0.21	0.50	1.00	1.00	1.00

Next, we identified the ERP amplitude features showing significant differences between the target and non-target, using two-sample t-tests with multiple comparison correction by FDR, where non-target feature values were derived from averaging the features of three non-targets. Then, we constructed a time-channel map of the t-values corresponding to those features that showed significant differences ($p < 0.05$) for each stimulus type (Fig. 6c). When comparing the average of absolute t-values among the stimulus types (beep: 3.533 ± 0.436, voice: 4.003 ± 0.896, animal: 3.524 ± 0.763), significant differences were found with ANOVA followed by a post-hoc analysis between the beep and voice ($p < 0.001$) and between the animal sounds and voice ($p < 0.001$), while there was no significant difference between the beep and animal sounds ($p > 0.05$). However, the number of features with a larger difference (FDR-adjusted p values < 0.01) were inspected (beep: 120, voice: 971, animal:

2004), showing the smallest with the beep. The time-channel t-value map also illustrated that the significant features appeared to be temporally localized, particularly with the beep and voice sounds. To associate this observation with BCI performance, we simulated classification accuracy offline using the portions of features within a particular time window: 0–200 ms, 200–400 ms, 400–600 ms (Fig. 6d). For the BCI with the beeps, the average accuracy derived by using the features between 200 and 400 ms was slightly higher than the online BCI accuracy using all the features. In contrast, for the BCIs with natural sounds, the online BCI accuracy was higher than those from any other windows.

Comparison between visual and auditory BCIs

We compared visual and auditory P300-based BCIs on the second day in the same way as we did for the three auditory BCIs on the first day, in terms of online BCI performance, P300 component characteristics, and the degree of feature differences between target and non-target.

The average (± standard deviation) classification accuracy of auditory and visual BCIs on the second day was 54.67% (± 23.45) and 77.56% (± 23.06), respectively (Fig. 3a). A paired t-test showed a significant difference in accuracy between two BCIs ($t(29) = -6.13$, $p < 0.001$). When the number of stimulus repetitions was increased from 1 to 7, a 175% increase in the classification accuracy was shown in the visual paradigms, while there was a 161% increase in the auditory one (Fig. 3b). In addition, the average ITR of the visual BCI systems was significantly higher than that of the auditory BCI system as the mean (± standard deviation) of ITR from auditory and visual BCI systems were 1.90 bits/min (± 2.09) and 4.47 bits/min (± 2.73), respectively ($t(29) = -5.97$, $p < 0.001$, paired t-test) (Fig. 3c).

The grand averaged ERPs induced by the target and non-target with visual and auditory stimuli on the second day are depicted in Fig. 5b, showing larger ERP amplitudes induced by the visual target stimuli than auditory target stimuli. Paired t-tests on the peak amplitude of P300 components between two modalities revealed significant differences at 19 channels, especially in frontal areas (Online Resource 1). In contrast, there were only two channels (FC5, CP5) that showed a significant difference in the peak latency of the P300 component between the two modalities (FC5: $p < 0.001$, CP5: $p < 0.05$).

An additional investigation regarding the differences between the amplitudes obtained by the target and the averaged ones induced by three non-targets was conducted. The difference waveforms and topographies are depicted in Fig. 7a, b, respectively. Paired t-test showed significant differences at 14 channels (FPz, Fz, FC1, FC2, CP2, P7, P8, Oz: $p < 0.05$, FC6, O1: $p < 0.005$, FT9, C3, Cz, O2: $p < 0.001$) between 200 and 800 ms ssof the ERP difference waveforms (Online Resource 1). In addition, topographies also showed differences between target and non-target in central, parietal, and occipital areas. Especially, while the visual stimuli induced large differences as early as 200 ms after stimulus onset, the auditory stimuli induced large differences around 400 ms after stimulus onset.

There were also dissimilarities between the two modalities in the time-channel maps of t-values corresponding to the features that showed significant differences between the target and non-target (Fig. 7c). The number of features with a larger difference (FDR-adjusted p-values < 0.01) was smaller with the auditory stimuli than with the visual stimuli (auditory: 1752, visual: 1959). In addition, with the auditory stimuli, negative t-values were mostly distributed in the frontal channels and positive t-values in the parietal and occipital areas. In contrast, negative and positive t-values were distributed more dynamically over the different areas with the visual stimuli. Especially, the spatial distributions of t-values appeared to be more synchronized for the visual stimuli; temporally synchronized positive and negative t-values appeared several times with the visual stimuli at approximately 200 ms, 250–500 ms, and 500–700 ms, which was less apparent with the auditory stimuli. Lastly, using the subset of features within the time windows of 0–200 ms, 200–400 ms, and 400–600 ms yielded lower accuracy than the online BCI accuracy for both visual and auditory BCIs (Fig. 7d).

Discussion

Reactive BCIs using auditory stimuli can provide an alternative means for the users with unreliable eye movements caused by severe disabilities or for those with limited visual processing in AR or VR environments. Previous studies have proposed the stimulation paradigms for auditory BCIs that present supportive visual guides or spatially localized auditory stimuli. However, a vision-free auditory BCI with no spatial information via a single auditory channel can simplify the BCI design and broaden the applications of BCIs. This study aimed to design auditory stimuli suitable for such a single-channel vision-free auditory BCI by exploring natural and artificial stimuli. In line with previous reports, the online BCIs with natural sounds showed better performance than those with artificial sounds in every subject.

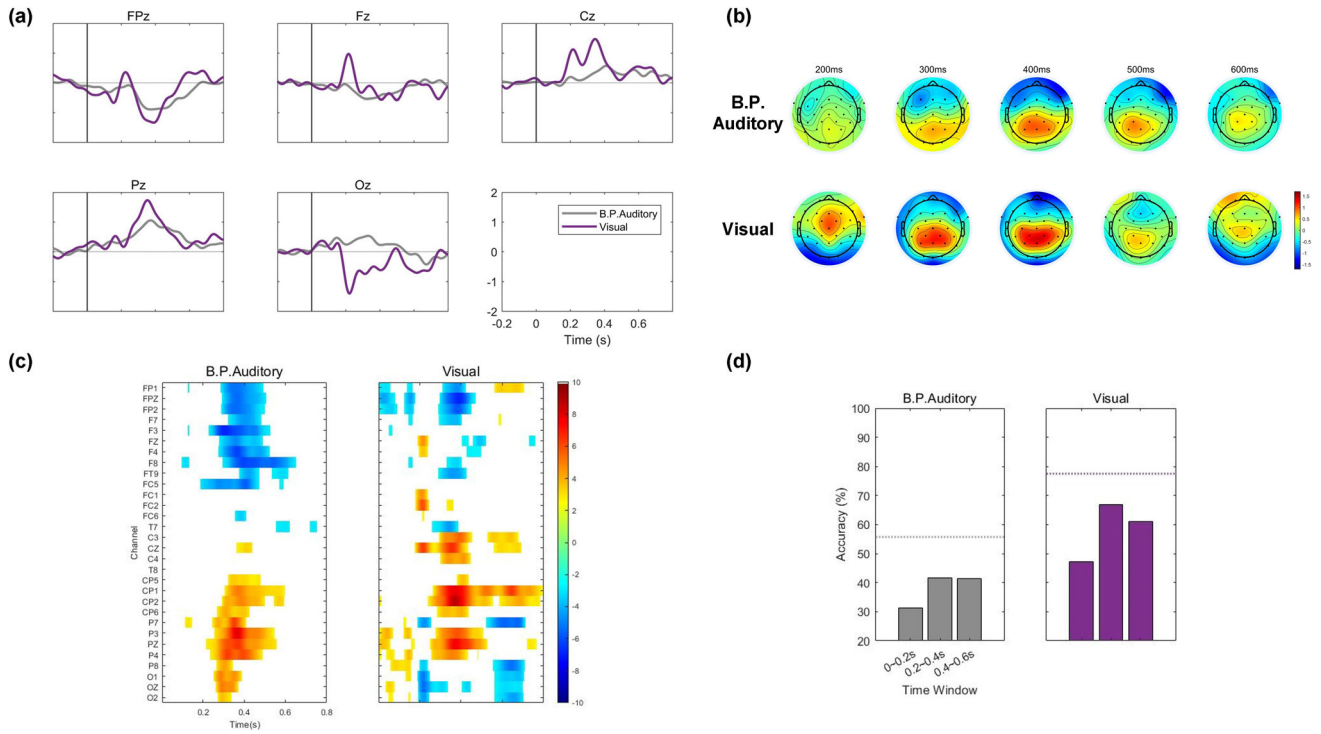

Fig. 7 **a** Grand averaged waveforms and **b** topographies showing the differences in ERP amplitudes of the target and nontarget induced by visual and auditory BCI systems. **c** Time-channel maps with t-values of each feature determined by two-sample t-tests, which were conducted to identify the features that showed significant differences between the target and nontarget. The features with positive t-values represent that the ERPs induced by the nontarget were bigger than the ones of the targets and they are colored in yellow or red. **d** Averaged classification accuracy derived by using the data of each different time window in each condition with the horizontal dotted lines representing the mean accuracy of the online tests

However, when we compared two natural sounds such as human voice and animal sounds, no specific type was better than the other across all the subjects; each subject exhibited the best performance with their preferred type of natural sounds. Moreover, in each subject, the BCI performance gap between the two natural sounds was bigger than that between the natural sound with lower performance and the artificial sound (beep). From this result, when designing an optimal sound for single-channel auditory BCIs in individuals, instead of simply choosing a natural sound over an artificial one, further exploration of a particular natural sound preferred by individuals is desired. Moreover, the search for individually optimized auditory stimuli among various types needs to be conducted based on online BCI operation, rather than other perceptual or survey-based procedures. This is partly supported by our observation of the discrepancy between perceptual performance and BCI performance. We calculated correlations between the accuracy of target recognition in the pre-training session and the accuracy of BCI control in the online BCI session and found low correlations: − 0.1759, 0.2811, and − 0.1146 for the beeps, voice, and animal sounds, respectively. Moreover, the perceptual performance difference between the two natural sounds was smaller (13.95%) than that between the natural sound with lower performance and the beep (28.20%). These results demonstrated that the performance in the pre-training session could not predict online BCI performance. In addition, the post-survey results showed that only 18 out of 30 subjects responded to the same stimuli as the most suitable for real-life device control as the one with the highest accuracy in their online BCI control.

We compared three types of auditory stimuli, including beeps, voices, and animal sounds. First, we identified the differences in the peak amplitude of P300 components among the three types. These differences were found at 10 channels, including Fz and Cz, where strong P300 responses are usually elicited, and T7 and CP5, where the left temporal area is related to rapid temporal information processing of both verbal and nonverbal auditory stimuli (Katayama and Polich 1996; Zaehle et al. 2004). In contrast, no pronounced difference was found in the peak latency. This was contrary to what we had expected as some subjects gave verbal feedback that the length of the stimuli was not perceived equal among the stimulus types even though they were all the same; they reported that the beeps were perceived as the shortest and the animal sounds were the longest. This may imply that the peak latency of

P300 depends on the actual length of the stimuli, not the perceived length.

Although using artificial sounds as auditory stimuli for P300-based BCIs resulted in the lowest performance, it may not rule out the possibility of exploring artificial sounds because the design with artificial sounds would be more straightforward and flexible. The analysis of differences in ERP features between the target and non-target revealed that the number of features with large differences was relatively small with the beeps. Moreover, these features were mostly observed between 200 and 400 ms after the stimulus onset. These relatively smaller differences might be related to low performance using the beeps. One possible reason for such small differences would be that the serial presentation of different beeps could unexpectedly create a stream of notes, like melody, to which the users could be easily oriented and thus struggled with ignoring non-target stimuli. Applying the beat and rhythm to the design of beeps may help the users to focus more easily on a target stimulus in a stream of different beeps (Schmidt-Kassow et al. 2016), which will be an interesting topic for further investigation.

The differences in ERP features between targets and non-targets appeared to be dependent on sensory modality. Using visual stimuli, the difference waveform was similar to the ERP waveform of the target itself (Fig. 7a), indicating relatively larger ERP deflections by target stimuli compared to non-target ones. On the other hand, using auditory stimuli, the difference waveform was arc-shaped (Fig. 6a), indicating smaller differences in ERP waveforms between target and nontarget. The time-channel t-value maps of significantly different features between target and non-target also exhibited dissimilar patterns between visual and auditory stimuli. The t-value map with the auditory stimuli showed a single pattern with the opposite polarity between the anterior and posterior areas and less clear temporal alignment. On the contrary, the t-value map with the visual stimuli showed multiple segregated patterns with the time-varying polarity between the anterior and posterior area, and more precise temporal alignment (Fig. 7c). The t-value map with the visual stimuli also displayed more dynamic patterns than that with the auditory stimuli, presumably reflecting more distinct spatiotemporal ERP patterns in response to target stimuli compared to those to non-target stimuli. This indicates that cortical processing of the target and non-target stimuli might be different between visual and auditory stimuli, at least in the stimulation paradigm employed in this study. Referring to the observation that the difference waveform was similar to the target response waveform when using visual stimuli, the design of auditory stimuli should consider enhanced suppression of responses to non-target, so that enhanced selective attention to target can lead to an increase in differences of ERPs between target and non-target, thus increasing the performance of the auditory BCI system.

Among the three types of auditory stimuli, the one resulting in the best BCI performance on the first-day experiment was always natural sounds and selected as an auditory stimulus type on the second-day experiment. However, the performance with the selected auditory stimulus type on the second day (54.67%) was worse than that on the first day with the individually best auditory stimuli 64.22% ($p < 0.05$, paired t-test). Twenty subjects exhibited decreased accuracy on the second day, among whom seven subjects showed the best performance with the human voice and seven subjects conducted pre-training tasks on the second day. It indicates that the decrease in accuracy on the second day was not coupled with a specific auditory stimulus type and that additional pretraining on the second day did not lead to an increase in the performance. Moreover, a correlation between the interval between the days of the two experiments and the decrease in accuracy was not significant (0.105). Although it is still elusive whether such a decrease in performance of auditory BCIs is simply incidental or signs an unknown effect of the repeated use of auditory BCIs, it may indicate that other factors such as the motivation of the users should be considered in the use of BCIs over multiple days (Baykara et al. 2016).

In our study, the auditory stimuli were presented through a single channel using earphones, considering cases where using multiple auditory channels for communication or device control is limited. Thus, this study designed auditory BCI systems with no spatial auditory information. However, some previous studies used headphones to present spatially localized stimuli for auditory BCIs (Belitski et al. 2011; Höhne et al. 2011; Ferracuti et al. 2013; Simon et al. 2015; Baykara et al. 2016). In addition, the behavioral study by Belitski et al. demonstrated that the error rates were lower when the stimuli were presented from spatially distributed locations, compared to the condition when the stimuli were presented from a single location (Belitski et al. 2011). Schreuder et al. (2010) also showed that the BCI performance was higher with the stimuli presented from different speakers, compared to the one with the stimuli all played by a single speaker. Although the present study explores individual optimal auditory stimuli without using multiple locations, if spatial information can additionally be provided to the subjects, higher BCI performance is expected with the enhancement of selective attention. Our follow-up studies will pursue the development of auditory BCIs embedding such spatial information.

In sum, this study investigated the design of auditory stimuli for P300-based BCIs and revealed that the selection of natural sounds should be optimized for individual users. As there has been no study that considered individual

preference of auditory stimuli, this study is the first that demonstrated that individually preferred auditory stimuli should be considered to design a P300-based auditory BCI system. It also showed differences in ERP waveforms between visual and auditory stimuli, particularly in the context of the spatiotemporal dynamics of ERPs. Even though the stimuli were presented in an identical way with the same stimulus duration and same inter-stimulus interval, these marked differences in ERP patterns between two modalities were found. These new findings may imply that the characteristics of ERP responses induced with the auditory BCI systems should be examined differently from those with the visual BCI system, such as developing new stimulation paradigms to provoke more dynamical spatiotemporal ERP patterns. We anticipate that ongoing efforts to improve auditory BCIs to be on par with visual BCIs will broaden the opportunities to apply BCIs to our daily life especially when visual processing is limited.

Author contributions Conceptualization: Y-JC and S-PK; methodology: Y-JC and S-PK; experiment: Y-JC; analysis: Y-JC; supervision: S-PK and O-SK; writing—original draft: Y-JC; writing—review and editing: S-PK and O-SK. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by the Korea Government (MSIT) (2017-0-00432, Development of non-invasive integrated CI SW platform to control home appliances and external devices by user's thought via AR/VR interface).

Data availability The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent The studies involving human participants were reviewed and approved by the Ulsan National Institute of Science and Technology, Institutional Review Board (UNIST-IRB-21-22-A). The participants provided their written informed consent to participate in this study.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s11571-022-09901-3.

References

Baykara E, Ruf CA, Fioravanti C et al (2016) Effects of training and motivation on auditory P300 brain-computer interface performance. Clin Neurophysiol 127:379–387. https://doi.org/10.1016/j.clinph.2015.04.054

Belitski A, Farquhar J, Desain P (2011) P300 audio-visual speller. J Neural Eng. https://doi.org/10.1088/1741-2560/8/2/025022

Bigdely-Shamlo N, Mullen T, Kothe C et al (2015) The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front Neuroinform. https://doi.org/10.3389/fninf.2015.00016

Birbaumer N, Cohen LG (2007) Brain-computer interfaces: communication and restoration of movement in paralysis. J Physiol 579:621–636. https://doi.org/10.1113/jphysiol.2006.125633

Carabalona R, Grossi F, Tessadri A et al (2010) Home smart home: Brain-computer interface control for real smart home environments. In: Proceedings of the 4th international convention on rehabilitation engineering & assistive technology. Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, p 51

Chang M, Nishikawa N, Struzik ZR et al (2013) Comparison of P300 responses in auditory, visual and audiovisual spatial speller BCI paradigms. arXiv preprint http://arxiv.org/abs/1301.6360

Chang CY, Hsu SH, Pion-Tonachini L et al (2020) Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings. IEEE Trans Bio-Med Eng 67:1114–1121. https://doi.org/10.1109/Tbme.2019.2930186

Cheng SY, Hsu HT, Shu CM (2008) Effects of control button arrangements on human response to auditory and visual signals. J Loss Prevent Proc 21:299–306. https://doi.org/10.1016/j.jlp.2007.03.002

Corralejo R, Nicolás-Alonso LF, Álvarez D et al (2014) A P300-based brain–computer interface aimed at operating electronic devices at home for severely disabled people. Med Biol Eng Compu 52:861–872

De Vos M, Kroesen M, Emkes R et al (2014) P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. J Neural Eng 11:036008. https://doi.org/10.1088/1741-2560/11/3/036008

Donchin E, Ritter W, McCallum WC (1978) Cognitive psychophysiology: the endogenous components of the ERP. Event-Relat Brain Potentials Man 349:411

Farwell LA, Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 70:510–523. https://doi.org/10.1016/0013-4694(88)90149-6

Ferracuti F, Freddi A, Iarlori S et al (2013) Auditory paradigm for a P300 BCI system using spatial hearing. In: IEEE international conference on intelligent robots, pp 871–876

Furdea A, Halder S, Krusienski DJ et al (2009) An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology 46:617–625. https://doi.org/10.1111/j.1469-8986.2008.00783.x

Halder S, Rea M, Andreoni R et al (2010) An auditory oddball brain-computer interface for binary choices. Clin Neurophysiol 121:516–523. https://doi.org/10.1016/j.clinph.2009.11.087

Halder S, Kathner I, Kubler A (2016) Training leads to increased auditory brain-computer interface performance of end-users with motor impairments. Clin Neurophysiol 127:1288–1296. https://doi.org/10.1016/j.clinph.2015.08.007

Harvey DG, Torack RM, Rosenbaum HE (1979) Amyotrophic lateral sclerosis with ophthalmoplegia—clinicopathologic study. Arch Neurol-Chicago 36:615–617. https://doi.org/10.1001/archneur.1979.00500460049005

Hayashi H, Kato S (1989) Total manifestations of amyotrophic lateral sclerosis—ALS in the totally locked-in state. J Neurol Sci 93:19–35. https://doi.org/10.1016/0022-510x(89)90158-5

Höhne J, Schreuder M, Blankertz B et al (2011) A Novel 9-class auditory ERP paradigm driving a predictive text entry system. Front Neurosci 5:99. https://doi.org/10.3389/fnins.2011.00099

Höhne J, Krenzlin K, Dahne S et al (2012) Natural stimuli improve auditory BCIs with respect to ergonomics and performance. J Neural Eng. https://doi.org/10.1088/1741-2560/9/4/045003

Huang MQ, Jin J, Zhang Y et al (2018) Usage of drip drops as stimuli in an auditory P300 BCI paradigm. Cogn Neurodyn 12:85–94. https://doi.org/10.1007/s11571-017-9456-y

Katayama J, Polich J (1996) P300 from one-, two-, and three-stimulus auditory paradigms. Int J Psychophysiol 23:33–40. https://doi.org/10.1016/0167-8760(96)00030-X

Kim M, Kim MK, Hwang M et al (2019) Online Home appliance control using EEG-based brain–computer interfaces. Electronics. https://doi.org/10.3390/electronics8101101

Klobassa DS, Vaughan TM, Brunner P et al (2009) Toward a high-throughput auditory P300-based brain-computer interface. Clin Neurophysiol 120:1252–1261. https://doi.org/10.1016/j.clinph.2009.04.019

Kübler A, Birbaumer N (2008) Brain-computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol 119:2658–2666. https://doi.org/10.1016/j.clinph.2008.06.019

Mullen TR, Kothe CAE, Chi YM et al (2015) Real-time neuroimaging and cognitive monitoring using wearable Dry EEG. IEEE Trans Bio-Med Eng 62:2553–2567. https://doi.org/10.1109/Tbme.2015.2481482

Ng AWY, Chan AHS (2012) Finger response times to visual, auditory and tactile modality stimuli. Lect Notes Eng Comput 2:1449–1454

Nicolas-Alonso LF, Gomez-Gil J (2012) Brain computer interfaces, a review. Sensors 12:1211–1279. https://doi.org/10.3390/s120201211

Oralhan Z (2019) A new paradigm for region-based P300 speller in brain computer interface. IEEE Access 7:106617–106626. https://doi.org/10.1109/Access.2019.2933049

Sara G, Gordon E, Kraiuhin C et al (1994) The P300 ERP component: an index of cognitive dysfunction in depression? J Affect Disord 31:29–38. https://doi.org/10.1016/0165-0327(94)90124-4

Schmidt-Kassow M, Wilkinson D, Denby E et al (2016) Synchronised vestibular signals increase the P300 event-related potential elicited by auditory oddballs. Brain Res 1648:224–231. https://doi.org/10.1016/j.brainres.2016.07.019

Schreuder M, Blankertz B, Tangermann M (2010) A New auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLoS ONE. https://doi.org/10.1371/journal.pone.0009813

Si-Mohammed H, Petit J, Jeunet C et al (2020) Towards BCI-based interfaces for augmented reality: feasibility, design and evaluation. IEEE Trans vis Comput Graph 26:1608–1621. https://doi.org/10.1109/Tvcg.2018.2873737

Simon N, Kathner I, Ruf CA et al (2015) An auditory multiclass brain-computer interface with natural stimuli: usability evaluation with healthy participants and a motor impaired end user. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2014.01039

Takano K, Hata N, Kansaku K (2011) Towards intelligent environments: an augmented reality-brain-machine interface operated with a see-through head-mount display. Front Neurosci. https://doi.org/10.3389/fnins.2011.00060

Wolpaw JR, Birbaumer N, McFarland DJ et al (2002) Brain-computer interfaces for communication and control. Clin Neurophysiol 113:767–791. https://doi.org/10.1016/s1388-2457(02)0005

Zaehle T, Wustenberg T, Meyer M et al (2004) Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur J Neurosci 20:2447–2456. https://doi.org/10.1111/j.1460-9568.2004.03687.x

Zander TO, Kothe C (2011) Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng 8:025005. https://doi.org/10.1088/1741-2560/8/2/025005

Zeng H, Wang YX, Wu CC et al (2017) Closed-loop hybrid gaze brain-machine interface based robotic arm control with augmented reality feedback. Front Neurorobotics. https://doi.org/10.3389/fnbot.2017.00060