A Low-Cost Auditory Multi-Class Brain Computer Interface based on Pitch, Spatial and Timbre Cues

P300-based brain-computer interfaces (BCIs) are especially useful for people with illnesses, which prevent them from communicating in a normal way (e.g. brain or spinal cord injury). However, most of the existing P300-based BCI systems use visual stimulation which may not be suitable for patients with sight deterioration (e.g. patients suffering from amyotrophic lateral sclerosis). Moreover, P300-based BCI systems rely on expensive equipment, which greatly limits their use outside the clinical environment. Therefore, we propose a multi-class BCI system based solely on auditory stimuli, which makes use of low-cost EEG technology. We explored different combinations of timbre, pitch and spatial auditory stimuli (TimPiSp: timbre-pitch-spatial, TimSp: timbre-spatial, and Timb: timbre-only) and three inter-stimulus intervals (150ms, 175ms and 300ms), and evaluated our system by conducting an oddball task on 7 healthy subjects. This is the first study in which these 3 auditory cues are compared. After averaging several repetitions in the 175ms inter-stimulus interval, we obtained average selection accuracies of 97.14%, 91.43%, and 88.57% for modalities TimPiSp, TimSp, and Timb, respectively. Best subject's accuracy was 100% in all modalities and inter-stimulus intervals. Average information transfer rate for the 150ms inter-stimulus interval in the TimPiSp modality was 14.85 bits/min. Best subject's information transfer rate was 39.96 bits/min for 175ms Timbre condition. Based on the TimPiSp modality, an auditory P300 speller was implemented and evaluated by asking users to type a 12-characters-long phrase. Six out of 7 users completed the task. The average spelling speed was 0.56 chars/min and best subject's performance was 0.84 chars/min. The obtained results show that the proposed auditory BCI is successful with healthy subjects and may constitute the basis for future implementations of more practical and affordable auditory P300-based BCI systems.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Author 1: Zacharias Vamvakousis *(Corresponding author)* / *Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain / email: zacharias.vamvakousis@upf.edu.*

Author 2: Rafael Ramirez */ Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

1. Introduction

1.1. Motivation

Brain-computer interfaces (BCIs) aim to develop computer systems capable of decoding useful information directly from brain activity in real-time (see (Wolpaw, 2000) for a review). Their objective is to enable direct communication between the brain and computers, with potential applications ranging from medicine to general consumer electronics. Over the past two decades BCI research has explored a variety of approaches for collecting, analyzing, and interacting with brain activity data. In most cases, the information is encoded voluntarily by the user, either by performing some mental task producing a measurable signal to be used as a command, or by selectively attending to one of the presented stimuli to encode a choice. Selective attention is often detected by observing event related potentials (ERPs), in particular the P300 wave whose occurrence is related to the person's reaction to a particular stimulus, and not to the physical attributes of the stimulus. P300 potentials, when recorded by electroencephalography (EEG), can be observed as a positive deflection in voltage with a latency (i.e. delay between the stimulus and the response) of roughly 250-500 milliseconds. They are usually elicited using the oddball paradigm, in which low-probability target stimuli are randomly mixed with high-probability non-target ones.

One of the obvious applications of P300-based BCIs is as a communication system for people who suffer from severe motor disabilities (e.g. brain or spinal cord injury), which prevent them from communicating in a normal way. However, most of the existing P300-based BCI systems rely on visual stimulation, which may not be suitable for patients with sight deterioration, such as patients suffering from Amyotrophic Lateral Sclerosis (ALS). In the case of patients who are unable to direct their gaze, adjust their focus or blink, an auditory P300-based interface might be a better alternative [2-9]. Furthermore, the use of auditory P300-based interfaces for patients with residual vision could allow visual stimuli to be used only as a feedback channel, therefore preventing interaction stimulation and feedback.

A second issue, if one wishes to improve the accessibility to BCI systems, and P300-based BCI systems in particular, is to reduce their cost. A limitation of P300-based systems is that they typically rely on expensive equipment with prices in the order of 30,000 USD or more, and are confined to experimental laboratories, which can be intimidating to some patients such as children and adults with cognitive disorders. In addition, setting up the BCI system at the beginning of each session can take an experienced clinical professional up to an hour to place the electrodes on the patient's scalp, which results in long and tedious sessions. Furthermore, typically such P300-based systems require the application of conductive gel in order to create a reliable connection between each electrode and the patient's scalp. The gel attaches to the patient's hair and can only be properly removed by washing the entire head at the end of each session. Recently, a number of low-cost EEG systems have been commercialized [28, 29]. They are mainly marketed as gaming devices and provide a limited solution to the expensive equipment problems described above: they are wirelessly connected to an ordinary computer, they require a short set-up time to adjust the electrodes to the user's scalp, and they do not require conductive gel. Recent research on evaluating the reliability of some of these low-cost EEG devices for research purposes has suggested that they are reliable for measuring visual and auditory evoked potentials [Duvinage, 2013; Debener, 2012; Badcock, 2013].

In this study, we propose a low-cost multi-class BCI system based solely on auditory stimuli. We explore different combinations of timbre, pitch and spatial auditory stimuli (TimPiSp: timbre-pitch-spatial, TimSp: timbre-spatial, and Timb: timbre-only) and three Inter-Stimulus intervals (150ms, 175ms and 300ms), and evaluate our system by conducting an oddball task on 7 healthy subjects. Additionally an auditory P300 speller is implemented and evaluated by asking users to type a phrase containing 12 characters.

1.2 Related work

P300 potentials can be observed as a positive deflection in voltage with a latency of roughly 250-500 ms with respect to an event [16,17]. Normally, P300 potentials are triggered by an attended rare event, so they are typically elicited using the oddball paradigm, in which low-probability target stimuli are mixed with high-probability non-target ones. In the past, visual P300 responses have been widely investigated for implementing BCIs [e.g. 13,14], and in particular for creating speller applications [15,19–21]. Similarly, auditory P300 responses have been used for implementing speller applications, e.g. [22]. In this study , a matrix of characters is presented for reference purposes with its columns and rows marked by a spoken number that is presented to the subject. Subjects are instructed to attend to the spoken number, which identifies the character. When the spoken number corresponding to the row or column containing the character is produced, it elicits a P300 wave, which can be detected from the EEG. The selected letter is identified according to the row and column that give a P300 response. The evaluation of the system produced satisfactory results with performance reaching up to 100% for one subject. However, it is clear that auditory stimulation with spoken numbers is time consuming, reducing the information transfer rate (selection of a letter can take 3.6 minutes).

In a more recent study [6], the spoken numbers were replaced by 6 natural sounds, which were mapped to rows and columns in an intuitive way allowing subjects to learn the mapping within a couple of sessions. Subjects were divided into two groups: one group was given auditory and visual stimulations while the other received only auditory stimulation. Although at the beginning of the experiment the accuracy of the auditory-only group was lower than the accuracy of the auditory-visual group, after 11 sessions their accuracy increased comparable to the the one of the auditory-visual group. Inter-Stimulus interval was 500 ms and the reported average ITR for the auditory modality was 1.86 bits/min.

Most oddball experiments use acoustic cues such as pitch, amplitude or length. However, other sound properties, such as spatial location of the stimulus, have been investigated. Teder-Sälejärvi et al. [12], conducted an oddball experiment in which an array of seven speakers (with a separation among them of 9 degrees) presented targets and non-targets in random order. Subject's attention to a particular direction elicited P300 responses. Another study [23], explored the use of virtual spatial localization to separate targets from non-targets through stereo headphones. Non-targets were produced from a straight direction (i.e. zero degrees) while targets were produced from a 30 and 90 degrees direction. The focus of this study was on early mismatch negativity potentials and not in P300 responses, engaging the subjects in passive listening while they were watching a film. A similar study [24] was conducted using free-field speakers with 10 degrees spatial separation.

In a more related study [7], a multi class BCI experiment, which used spatially distributed, auditory cues was conducted. The stimulus set consisted of 8 stimuli, different in pitch. The subjects were surrounded by 8 free field speakers, each of which was assigned to one of the stimuli. In the experiment, 10 subjects participated in an offline oddball task with the spatial location of the stimuli being a discriminating cue. The experiment was conducted in free field, with an individual speaker for each location. Different inter-stimulus intervals were investigated: 1000, 300, and 175 ms. Average accuracies were over 90% for most conditions, with corresponding information transfer rates up to an average of 17.39 bits/minute for the 175 ms condition (best subject 25.20 bits/minute). Interestingly, when discarding the spatial cues by presenting the stimuli through a single speaker, selection accuracies dropped below 70% for most subjects.

In a later study [8], the same authors implemented an auditory speller using the same stimuli presentation design, but reducing the set to 6 sounds. In order to optimize the spelling speed, a dynamic stopping method was introduced. This method minimized the number of repetitions required for each trial. Sixteen out of 21 subjects managed to spell a sentence in the first session. These subjects were selected for a second session where they were asked to type two sentences. In the second session an average of 5.26bits/min (0.94char/min) ITR was achieved, which sets the current state of the art in auditory P300 spellers.

A very similar auditory BCI system using spatially distributed, auditory cues is proposed by Käthner et al. [9]. The set of free field speaker is replaced by stereo headphones. Different ISIs of 560, 400, 320, 240 and 160 ms were evaluated in a P300 auditory speller paradigm. An average of 2.76 bits/min was reported under the 400 ms ISI condition. Unfortunately the training of the classification process was performed only for the 560ms ISI. The acquired classifier was then used for all studied ISIs. This resulted to the conclusion that bigger ISIs give better selection accuracy. The opposite results were obtained by Schreuder et al. [7], when a separate classifier was trained for each condition.

Other researchers have investigated the feasibility of using the Emotiv EPOC device for detecting auditory ERPs. Badcock et al.. [1] simultaneously recorded, using research and Emotiv Epoc devices, the EEG of 21 subjects while they were presented with 566 standard (1000 Hz) and 100 deviant (1200 Hz) tones under passive and active conditions. For each subject, they calculated auditory ERPs (P1, N1, P2, N2, and P300 peaks) as well as mismatch negativity (MMN) in both active and passive listening conditions. They restricted their analysis to frontal electrodes. Their results show that the morphology of the research and Emotiv Epoc EEG system late auditory ERP waveforms were similar across all participants, but that the research and gaming EEG system MMN waveforms were only similar for participants with non-noisy MMN waveforms. Peak amplitude and latency measures revealed no significant differences between the size or the timing of the auditory P1, N1, P2, N2, P3, and MMN peaks. Based on these results they conclude that the Emotiv Epoc EEG system may be a valid alternative to research EEG systems for recording reliable auditory ERPs.

In another study [31], Emotiv Epoc was combined with a standard infracerebral electrode cap with Ag/AgCl electrodes. The result was a low-cost portable EEG system that was tested in an auditory oddball paradigm under sitting and walking conditions. With an ISI of 1 second, the single trial accuracy was 77% for sitting and 69% for walking conditions. In a later study [32] -using the same EEG system-, the conclusion that a low-cost single trial portable EEG interface is feasible is enforced.

2. Materials and methods

2.1 Participants

All subjects taking part in the present study gave written informed consent to be involved in the research and agreed to their anonymized data to be analyzed. Procedures were positively evaluated by the Parc de Salut MAR - Clinical Research Ethics Committee, Barcelona, Spain, under the reference number: 2013/5459/I. Seven healthy adults (3 female, 4 male, mean age 42 years) participated in a multi-class auditory oddball paradigm. Subjects reported to have normal hearing, and no difficulty with spatial localization of sounds in everyday situations.

2.2 Data Acquisition

The Emotiv EPOC EEG system [28] was used for acquiring the EEG data. It consists of 16 wet saline electrodes, providing 14 EEG channels, and a wireless amplifier (with a sample rate of 128 Hz). The 16 electrodes are aligned with positions in the 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, FC4, M1, and M2. The electrode positioned at M1 acts as reference electrode, while the electrode at M2 is used for reducing external electrical interferences. The EEG signals were sampled at 128 Hz, digitized with a resolution of 16 bits, and band-pass filter with a 4th order Butterworth 1-12Hz filter.
We collected and processed the data using the OpenViBE platform [10]. In order to trigger virtual instrument sounds through the OpenVibe platform, a VRPN to midi gateway was implemented and used along with LoopBe virtual MIDI port[1]. Sound stimulus was then played

1 "LoopBe1 - A Free Virtual MIDI Driver - Nerds.de." 2004. 11 Nov. 2013 <http://www.nerds.de/en/loopbe1.html>

back by Propellerhead Reason[2] virtual instrument host application. MBOX low-latency sound card was used, offering 17 ms output latency. The LoopBe MIDI port used introduced an additional latency of 1 to 3 ms. Both data acquisition and on-line scenario were performed on a laptop with an Intel Core i5 2,53 Ghz processor with 4 GB of RAM, running windows 7 64-bit Operating System.

2.3. Experiment Design

2.3.1 Auditory modality Experiment

In all sessions, subjects were asked to sit motionless in a comfortable chair facing two loudspeakers, Roland MA-150U placed at 45 and -45 degrees with respect to the subject's orientation. The speakers were placed 15cm below ear level and approximately at one meter from the subject (see Figure 1). The speakers were set to equal loudness intensity of approximately 60 dB for every stimulus. Subjects were initially exposed to each stimulus in isolation and then to the stimuli mix in order to familiarize them with the sounds. At the beginning of each experiment, subjects were asked to close their eyes, minimize their eye movements and avoid moving during the experiment. All the experiments were designed as an auditory oddball task. The room was not electromagnetically shielded, and no extensive sound attenuating precautions were taken.

Three different ISI were explored: 300 ms and 175 ms and 150 ms. For the 300 ms and 175 ms conditions three different stimuli discriminating cues were examined: timbre only (Timb), timbre and spatial (TimSp), and timbre, pitch and spatial (TimPiSp). For the 150ms condition only the TimPiSp modality was studied. In all conditions the stimulus set consisted of 6 short sounds (of a duration of 100ms). In total 7 different conditions were studied: TimPiSp-150ms ISI (TimPiSp150), TimPiSp-175ms ISI (TimPiSp175), TimPi-175ms ISI (TimPi175), Timb175 (Timb175), TimPiSp-300ms ISI (TimPiSp300), TimPi-300ms ISI (TimPi300), Timb175 (Timb300).

In the Timb conditions, all stimuli were generated with different timbre but with fixed pitch (130.81 Hz) and spatial location (center); in the TimSp conditions, stimuli were generated with different timbre and spatial location but fixed spatialization; and in TimPiSp conditions all timbre, pitch and spatialization were differentiated (see Table 1). Blocks of the different conditions were mixed to prevent time biases. For each condition, a training session was followed by an online session. This resulted in 14 sessions for every subject. The collected EEG data of each training session were used for acquiring a spatial filter and a Linear Discriminant Analysis Classifier, used in the on-line classification process. Both the training and the on-line sessions consisted of ten trials. In the 300ms condition each trial consisted of 90 sub-trials, 15 for each stimuli, while in the 175 and 150ms conditions each trial consisted of 150 sub-trials, 25

2 "Reason - Complete music making, music production ... - Propellerhead." 11 Nov. 2013 <http://www.propellerheads.se/products/reason/>

for each stimulus. This resulted in 900 sub-trials per session (150 of which target) in the 300ms condition and 1500 sub-trials per session (250 of which target) in the 175 and 150ms conditions.

Before each trial a random stimulus was selected as the target stimulus and was played back to the subject (see figure 2). A trial can be divided into N repetitions (where N is 15 for the 300ms conditions and 25 for the 175 and 150ms conditions). A repetition consists of a random sequence of all 6 stimuli. An example of a repetition's stimuli presentation for the TimPiSp175 condition is shown in figure 3. Stimuli were randomized in a way that the same stimulus never appeared consecutively. The subjects were instructed to tap on the desk every time the target stimulus appeared and mentally count its occurrences. In the on-line session, 1.9 seconds after each trial, the stimulus detected as target was played back to the subject followed by an interval of 3 seconds before presenting the target stimulus of the next trial.

2.3.2 Speller Experiment

In the speller experiment the subjects were asked to spell a 12-characters phrase in Spanish ("HOLA QUE TAL"). The speller experiment was very similar to the BCI experiment: speakers were positioned in the same way, the random sequence stimuli presentation was identical, and during a trial the subject was asked to keep their eyes closed. However, in the speller experiment only the TimPiSp150 and TimPiSp175 conditions was examined (depending on the performance of each user for each condition). At the beginning of each experiment, subjects were asked to become familiar with the speller interface, i.e. the mapping of stimuli into letters in the alphabet (see figure 6). Then while stimuli were played in a random order, subjects were asked to switch their attention to each of the 6 stimuli. This process lasted until each subject could quickly switch his or her attention to all 6 stimuli (about 10 minutes). The reason for that task was that in the case of the speller the target sound is not played back to the users before each trial, so the task of focusing on the target stimulus becomes more difficult.

Before the spelling session, one more training session -identical to the one described in the auditory modality experiment- was conducted in order to acquire the spatial filter and LDA classifier to be used in the spelling session. In the spelling session, the speller interface was used to select letters in two selection steps. First a group of letters was selected by selecting a column in the speller interface, i.e. by focusing attention on the stimulus corresponding to the column to be selected. In the second step, a particular letter was selected from the groups of letters previously selected, by focusing attention on the stimulus corresponding to the row containing the letter. One stimulus was reserved to specify the "undo" action used to return the subject to the first selection step (organ sound). In the case of a misspelled character, the users had to select the backspace character in order to delete it. The speller interface, the text to be written, and the subject's progress were presented visually. After each trial the detected sound stimulus was played back to the user and the user's progress was updated. Between the trials, six of the users were instructed orally on what the next target sound should be. One user was sufficiently familiar with the interface in order to complete the task without any oral instructions.

2.4 Analysis

2.4.1. Training Session

The recording of the training sessions were analyzed in order to acquire a spatial filter and a two class (target, non-target) LDA classifier (see figure 5). First, the signal was preprocessed by applying a band pass filter in the range of 1 to 12 Hz, and down-sampled to 32 Hz. Given the noisy nature of the EEG signal, a xDawn spatial filter was applied in order to enhance the P300 response. The xDAWN algorithm [26] allows the estimation of a set of spatial filters for optimizing the signal to signal-plus-noise (SSNR) ratio. The xDAWN method assumes that there exists a typical response synchronized with the target stimuli superimposed on an evoked response to all the stimuli, and that the evoked responses to target stimuli could be enhanced by spatial filtering. A window of 250 to 750 ms after the stimuli presentation was applied to train the xDAWN algorithm in order to acquire a 14 to 3 channels spatial filter. This resulted in a matrix of 48 features. No additional artifact rejection method was applied. All epochs were used in the training and classification process.

The features produced by the xDAWN filter were used to train a classifier of the form:

$$f(Fs([t+250,t+750])) \rightarrow \{target, non\text{-}target\}$$

where t is the stimulus presentation time, Fs([t+250,t+750]) is the feature set generated by the spatial filter, and target and non-target are the classes to be discriminated. Classification was performed by applying linear discriminant analysis (LDA) to the training data. LDA finds a linear combination of features, which separates two or more classes of objects or events. The resulting combination may be used as a linear classifier.

2.4.2 Online session

During the online session, the 48 features vector for each epoch were fed to the obtained LDA classifier (figure 6), whose output consisted of the vector distance to the hyper-plane (negative value for targets and positive for non-targets). These values were fed into a voting classifier. When the corresponding number of repetitions is reached, the voting classifier sums up the hyper-plane distances for all the repetitions of each stimuli. The stimulus with minimum sum is selected as the predicted target for that trial.

2.4.3 Information Transfer Rate

The information transfer rate (ITR) [27], i.e. the amount of information carried by every selection, can be computed as follows:

$$ITR(bits/\min) = S \cdot \left[\log_2(N) + P \cdot \log_2(P) + (1-P) \cdot \log_2\left(\frac{1-P}{N-1}\right)\right]$$

,where ITR is the number of bits per minute, S represents the number of selections per minute, N represents the number of possible targets, and P represents the probability that they are correctly classified. Note that increasing S by decreasing the number of repetitions would not necessarily increase the ITR because the accuracy of the classifier (i.e. P) will decrease. Thus, there is a tradeoff between S and P, and the choice of which is more important depends on the type of BCI application.

4. Results and Discussion

4.1. Auditory Modality Experiment

4.1.1. Accuracy and ITR

We distinguish between two accuracy measures: classification and selection accuracy. Classification accuracy refers to the percentage of sub-trials that is correctly identified as target or non-target. Selection accuracy refers to the percentage of trials in which the target stimulus is correctly identified. Given that we are interested in detecting target stimuli, in the following we report on selection accuracy.

In order to investigate the system's accuracy for different number of repetitions, the voting classifier object in OpenVibe platform was modified to keep a log of the hyper-plane distances' sums of each stimulus for any number of repetitions.

Tables 2,3 and 4 provide the online accuracy of all subjects and conditions along with the number of repetitions in the on-line sessions. Figure 7 shows the average accuracy and ITR (among subjects) for different number of repetitions. The ITR is considered to be zero, if the average accuracy is less than 70%.

The maximum accuracy is found in the TimPiSp175 condition (97.1%), followed by the TimPiSp150 (92.86%), TimbSp 175 (91.4%), Timb175 (88.57%), TimPiSp300 (88.57%), TimbSp300 (84.3%) and Timb300 condition (80%).

The average accuracy exceeds 70% in all conditions after 10 repetitions and 80% after 15 repetitions, while after around 18 repetitions the online accuracy does not improve significantly in all conditions (see figure 7). For a given number of repetitions, the 300 ms condition does not seem to provide better accuracy than the 175 ms and 150 ms conditions and as a result gives lower ITR. The maximum average ITR is achieved with around 10-15 repetitions for all conditions. In the TimPiSp175 condition the average accuracy is more than 90% after 19 repetitions.

The maximum average ITR is found In the TimPiSp150 condition (14.85 bits/min, with an average of 9.43 iterations). The best subject's performance was in the Timb175 condition (39.96 bits/min, accuracy 80% with 2 repetitions).

4.1.2 Physiological Response

For each condition and every subject, the training and on-line session EEG recordings were merged into one dataset and analyzed in Matlab using EEGlab [30] and ERP toolbox[3]. This resulted in 3000 sub-trials (500 targets) for the 300 ms modality and 1800 sub-trials (300 targets) for the 175 and 150 ms modalities, for each subject and condition. A window of 200 ms before the stimulus presentation was used for baseline removal. In all conditions a threshold of ±150μV was used for rejecting epochs with artifacts. The percentage of rejected epochs for each condition is shown in tables 1, 2 and 3. Since during the experiment, subjects remained still and with their eyes closed, the high artifact rejection rate between sessions (raging from 0% to 74.4% for the same user) is due to noise introduced by the Emotiv Epoch. Although the signal was always checked before every session, some EEG channels became noisy in the middle of a session.

Initially a grand average for all 7 conditions was created for each subject, and its P300 peak amplitude in the interval 250 and 650 ms was computed for all EEG channels for the target epochs. For each subject, the EEG channel with the highest P300 peak values was selected for further analysis. Tables 2, 3 and 4 show the averaged P300 amplitude and latency for all conditions and users. figures 8 and 9 show the averaged target and non-target responses of each user's selected channel for all the 175 and 300 ms ISI conditions, respectively. In all plots, the red line corresponds to target epochs and the black line to non-target epochs. A periodicity of 175 ms can be observed in the 175 ms condition and a periodicity of 300 ms in the case of 300 ms condition. As expected, this periodicity aligns with the stimuli presentation periodicity (see figure 3).

Figures 10, 11 and 12 show the average of all users' target and non-target responses for all 300ms ISI conditions of 10 EEG channels. When comparing the 3 modalities, it is observed that while the target ERP responses are equally strong in all modalities, the TimPiSp gives the weakest non-target ERP responses, followed by the TimPi and the Timb modalities. This results in a stronger mismatch negativity value. This is also reflected in the selection accuracies of each of these modalities: 88.5%, 84.3% and 80% for the TimPiSp, TimSp and Timb modality respectively.

4.2. Speller Results

Table 5 shows the results of the speller experiment. Six out of seven subjects completed the task. Best subject's ITR is 4.37 bits per minute, while average ITR was 3.04 bits/min. This resulted in an average spelling speed of 0.56 chars/min (best performance 0.84 chars/min). The non-linear correlation between the ITR and spelling speed is due to the fact the subjects should delete and retype the misspelled characters. The average on-line selection accuracy for the subjects that successfully completed the task was 82.45%. As predicted by Kübler et al.. [25], a selection accuracy of 70% is required for a useful BCI and all 6 subjects with an accuracy of more than 70% managed to spell all 12 characters, while one subject with accuracy 63.41% managed to

3 "ERPLAB Toolbox Home — ERP Info Home Page." 2008. 12 Nov. 2013 <http://erpinfo.org/erplab>

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

spell only 7 characters before abandoning the task after 44 minutes. Table 5 summarizes the results for all 6 subjects that completed the task.

4.3 Discussion

We propose a new experimental paradigm for a low-cost P300 based auditory BCI. For the first time the significance -in an auditory P300 paradigm- of the 3 most important perceptual auditory discriminating cues is studied: Timbre, Pitch and Spatialization, under three possible ISI conditions (300, 175 and 150 ms). The results of our study indicate that the best results are given when the stimuli are different in all three perceptual modalities, while shorter ISI results in higher ITR.

As seen in figures 8 and 9 all subjects have clear EPR responses in both the 175 and 300 ms conditions, although they vary in intensity and shape. The mean latency of the P300 peak for all 7 conditions is 468 ms, while no significant differences in the P300 peak amplitude and latencies are observed between the different conditions (see tables 2, 3, 4). Although the signal quality was checked at the beginning of each session, high epoch rejection rate was observed in some sessions. This might be due to the unstable behavior of saline water electrodes.

The channels with the strongest average P300 peak for all conditions were located in the frontal area for all subjects. When looking at the occipital channels though (figures 10, 11, 12), we can see an early positive deflection about 220 ms after the target stimuli presentation. This aligns with the results of Schreuder et al. [7], where it is concluded that in the short 175 ms condition "class difference has shifted toward the frontal areas when compared to the longer 1000 ms ISI condition".

Despite using a low-cost EEG device, the performance of the proposed system is comparable to state-of-the-art performance. In the TimPiSp150 condition the average selection accuracy obtained is 92.86% with 17.1 repetitions and the average ITR is 14.85 bits/min with 9.43 repetitions. These results compare well with the state-of-the-art results reported by Schreuder et al.. [7] (selection accuracy 94%, with 11.6 repetitions; maximum ITR of 17.39 bits/min, with 5.6 repetitions, PitchSpatial 175ms ISI). As it is seen in table 6, the average ITR achieved in the spelling paradigm is just below the state-of-the-art results, reported by Schreuder et al.. [8]. However, Schreuder et al.. use a dynamic stopping method used, which minimizes the number of repetitions per trial. The shorter ISI (150 ms), and the use of 3 auditory discriminating cues might have compensated the noisier signal acquired by a low-cost EEG system, resulting in a comparable ITR value.

The maximum average selection accuracy is found in the TimPiSp175 condition (97.1%), followed by the TimPiSp150 (92.86%), TimbSp175 (91.4%), Timb175 (88.57%), TimPiSp300 (88.57%), TimbSp300 (84.3%) and Timb300 condition (80%). The 300ms ISI conditions though were studied for a maximum of 15 repetitions, while the 175 and 150 ms ISI conditions were studied for a maximum of 25 repetitions. Looking at figure 7, we can see that for the same number of repetitions, the average accuracy is close for the 300 and 175 ms ISI conditions. The ITR though is much lower in the case of 300ms ISI conditions, as more time is required for the same number of repetitions. Thus, it is concluded that there is no reason for using long ISIs in auditory P300 based BCIs. In order to get a significantly stronger P300 response, When comparing the TimPiSp175 with TimPiSp150 conditions, we see that although the first one gives better selection accuracy (97.1% versus 92.86%), the second one achieves higher ITR (14.85 versus 10.1 bits/min). In the future, the ISI's limits should be studied in order to determine the minimum ISI to maximize ITR.

In both 300 and 175 ms ISI conditions, the order of the conditions in terms of selection accuracy is: TimPiSp, TimSp, Timb. Thus, it is clear that the performance of the system improves as more discriminating cues are added. This is also concluded when observing the averaged ERP responses of these conditions (figure 12). Although the target stimuli responses have the same intensity in all conditions, the non-target stimuli responses become weaker as more modalities are added in the stimuli design. This results in higher mismatch negativity values and thus, higher selection accuracy.

Schreuder et al.. emphasized the importance of sound spatialization in stimuli presentation. However, in their case stimuli differed only in pitch and spatialization. In their study, selection scores went down below 70% for most subjects when the spatialization modality was removed. Our results imply that when stimuli are different in timbre, the spatialization still affects the selection accuracy, but not so drastically. In the 300ms ISI conditions, the average accuracy of TimSp modality is 84.3% while in the Timbre modality the accuracy is 80%. In the 175 ms ISI conditions, the average accuracy of TimSp modality is 91.4% and the accuracy of the Timbre modality is 88.57%.

As seen in table 5, the online accuracy in the Speller experiment is 82.45%, while for the same conditions and subject the average in the auditory modality experiment was 96.67%. This lower performance of the speller, compared to the online performance, was also reported by Schreuder et al. [7, 8], where the average accuracy in the BCI experiment was 94%, while in the speller the average accuracy is 77.4%, resulting in a lower ITR. This difference can be explained by three reasons. Firstly, in the speller experiment, the target sound is not played back to the users, so the users have to memorize the sound stimulus. Secondly, the auditory speller consists of a much bigger amount of trials. This might lead to loss of concentration due to tiredness.

5. Conclusions

We have presented a multi-class BCI system based solely on auditory stimuli, which makes use of low-cost EEG technology. We have explored timbre-pitch-spatial, timbre-spatial, and timbre-only combinations of timbre, pitch and spatial auditory stimuli and three inter-stimuli intervals (150ms, 175ms and 300ms). We evaluated the system by conducting an oddball task on 7 healthy subjects. The maximum accuracy is found in the TimbPiSp175 condition (97.1%), followed by the TimPiSp150 condition (92.86%), TimbSp175 condition (91.4%), Timb175 condition (88.57%), TimPiSp300 condition (88.57%), TimbSp300 condition (84.3%) and Timb300 condition (80%). The maximum average ITR is found in the 150ms ISI, TimPiSp condition (14.85 bits/min, with 9.43 iterations). Lower Inter-Stimulus Intervals lead to higher ITR, while as more discriminating cues are added the selection accuracy and ITR increases. Based on the TimPiSp modality, an auditory P300 speller was implemented and evaluated by asking users to type a 12-characters-long phrase. Six out of 7 users completed the task. The average spelling speed was 0.56 chars/min and best subject's performance was 0.84 chars/min.

In this study we made use of an EEG device which is valued at about 50-100 times less costly than medical/research quality devices. However, interestingly our results are comparable to those achieved by medical devices. The obtained results show that the proposed auditory BCI is successful with healthy subjects and may constitute the basis for future implementations of more practical and affordable P300-based BCI systems. However, the high amount of noise introduced during some of the sessions (high epoch rejection rate in off-line analysis) affects the accuracy of the system, and thus for crucial BCI applications a more robust and stable EEG device should be used.

References

[1]Badcock, Nicholas A et al.. "Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs." *PeerJ* 1 (2013): e38.

[2] Hill, N Jeremy et al.. "An auditory paradigm for brain-computer interfaces." *Advances in Neural Information Processing Systems* 2004: 569-576.

[3]Kanoh, Shin'ichiro, Ko-ichiro Miyamoto, and Tatsuo Yoshinobu. "A brain-computer interface (BCI) system based on auditory stream segregation." *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* 20 Aug. 2008: 642-645.

[4] Kübler, Andrea et al.. "A Brain–Computer Interface Controlled Auditory Event-Related Potential (P300) Spelling System for Locked-In Patients." *Annals of the New York Academy of Sciences* 1157.1 (2009): 90-100.

[5] Furdea, A et al.. "An auditory oddball (P300) spelling system for brain-computer interfaces." *Psychophysiology* 46.3 (2009): 617-625.

[6] Klobassa, DS et al.. "Toward a high-throughput auditory P300-based brain–computer interface." *Clinical Neurophysiology* 120.7 (2009): 1252-1261.

[7] Schreuder, Martijn, Benjamin Blankertz, and Michael Tangermann. "A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue." *PLoS One* 5.4 (2010): e9813.

[8] Schreuder, Martijn, Thomas Rost, and Michael Tangermann. "Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI." *Frontiers in neuroscience* 5 (2011).

[9] Käthner, Ivo et al.. "A portable auditory P300 brain–computer interface with directional cues." *Clinical neurophysiology* (2012).

[10] Renard, Yann et al.. "OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments." *Presence: teleoperators and virtual environments* 19.1 (2010): 35-53.

[11] Nambu, Isao et al.. "Estimating the Intended Sound Direction of the User: Toward an Auditory Brain-Computer Interface Using Out-of-Head Sound Localization." *PloS one* 8.2 (2013): e57174.

[12] Teder-Sälejärvi, Wolfgang A, and Steven A Hillyard. "The gradient of spatial auditory attention in free field: an event-related potential study." *Perception & psychophysics* 60.7 (1998): 1228-1242.

[13] Bayliss, Jessica D, Samuel A Inverso, and Aleksey Tentler. "Changing the P300 brain computer interface." *CyberPsychology & Behavior* 7.6 (2004): 694-704.

[14] Piccione, Francesco et al.. "P300-based brain computer interface: reliability and performance in healthy and paralysed participants." *Clinical neurophysiology* 117.3 (2006): 531-537.

[15] Sellers, Eric W, and Emanuel Donchin. "A P300-based brain–computer interface: initial tests by ALS patients." *Clinical neurophysiology* 117.3 (2006): 538-548.

[16]Picton, Terence W. "The P300 wave of the human event-related potential." *Journal of clinical neurophysiology* 9.4 (1992): 456-479.

[17] Polich, John. "Updating P300: an integrative theory of P3a and P3b." *Clinical neurophysiology* 118.10 (2007): 2128-2148.

[18] Wolpaw, Jonathan R et al.. "Brain-computer interface technology: a review of the first international meeting." *IEEE transactions on rehabilitation engineering* 8.2 (2000): 164-173.

[19] Farwell, Lawrence Ashley, and Emanuel Donchin. "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials." *Electroencephalography and clinical Neurophysiology* 70.6 (1988): 510-523.

[20] Lenhardt, Alexander, Matthias Kaper, and Helge J Ritter. "An adaptive P300-based online brain–computer interface." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 16.2 (2008): 121-130.

[21] Nijboer, F et al.. "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis." *Clinical neurophysiology* 119.8 (2008): 1909-1916.

[22] Furdea, A et al.. "An auditory oddball (P300) spelling system for brain-computer interfaces." *Psychophysiology* 46.3 (2009): 617-625.

[23] Sonnadara, Ranil R, Claude Alain, and Laurel J Trainor. "Effects of spatial separation and stimulus probability on the event-related potentials elicited by occasional changes in sound location." *Brain research* 1071.1 (2006): 175-185.

[24] Deouell, Leon Y et al.. "Spatial location is accurately tracked by human auditory sensory memory: evidence from the mismatch negativity." *European Journal of Neuroscience* 24.5 (2006): 1488-1494.

[25] Kübler, Andrea et al.. "Predictability of Brain-Computer Communication." *Journal of Psychophysiology* 18.2-3 (2004): 121.

[26] Rivet, Bertrand et al.. "xDAWN algorithm to enhance evoked potentials: application to brain–computer interface." *Biomedical Engineering, IEEE Transactions on* 56.8 (2009): 2035-2043.

[27] Shannon, Claude E, and Warren Weaver. "The mathematical theory of communication (Urbana, IL." *University of Illinois Press* 19.7 (1949): 1.

[28] "Emotiv | EEG System | Electroencephalography." 2005. 11 Nov. 2013 <http://emotiv.com/>

[29] "NeuroSky - Brainwave Sensors for Everybody." 2005. 12 Nov. 2013 <http://www.neurosky.com/>

[30] Delorme, Arnaud, and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *Journal of neuroscience methods* 134.1 (2004): 9-21.

[31] Debener, Stefan et al.. "How about taking a low-cost, small, and wireless EEG for a walk?." *Psychophysiology* 49.11 (2012): 1617-1621.

[32] De Vos, Maarten, Katharina Gandras, and Stefan Debener. "Towards a truly mobile auditory brain–computer interface: Exploring the P300 to take away." *International Journal of Psychophysiology* (2013).

[33] Duvinage, Matthieu et al.. "Performance of the Emotiv Epoc headset for P300-based applications." Biomedical engineering online 12.1 (2013): 56.

Figure 1

Experiment setup

For all experiments two loudspeakers were used to spatialize the stimuli.

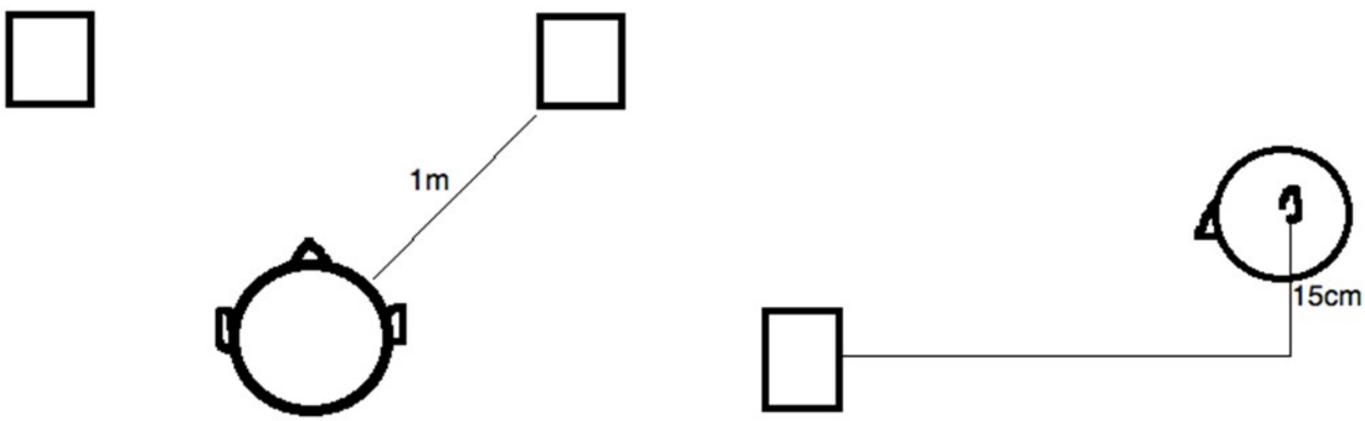


Figure 2

A session of the 175 ms ISI condition

Each session consisted of 10 trials. Before each trial, a random stimulus was played back as the target stimulus. In the case of 175ms ISI conditions a trial consisted of 25 repetitions of all stimuli in a random order and lasted for 26.25secs. In the case of 300ms (15 repetitions) and 150ms (25 repetitions) ISI conditions each trial lasted 27 and 22.5 seconds, respectively. In the on-line sessions, the detected target stimulus was played-back after each trial.

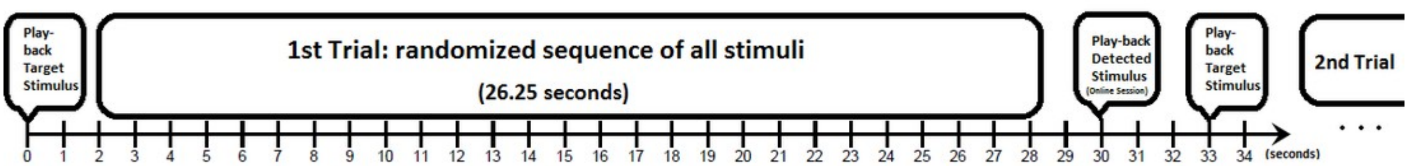
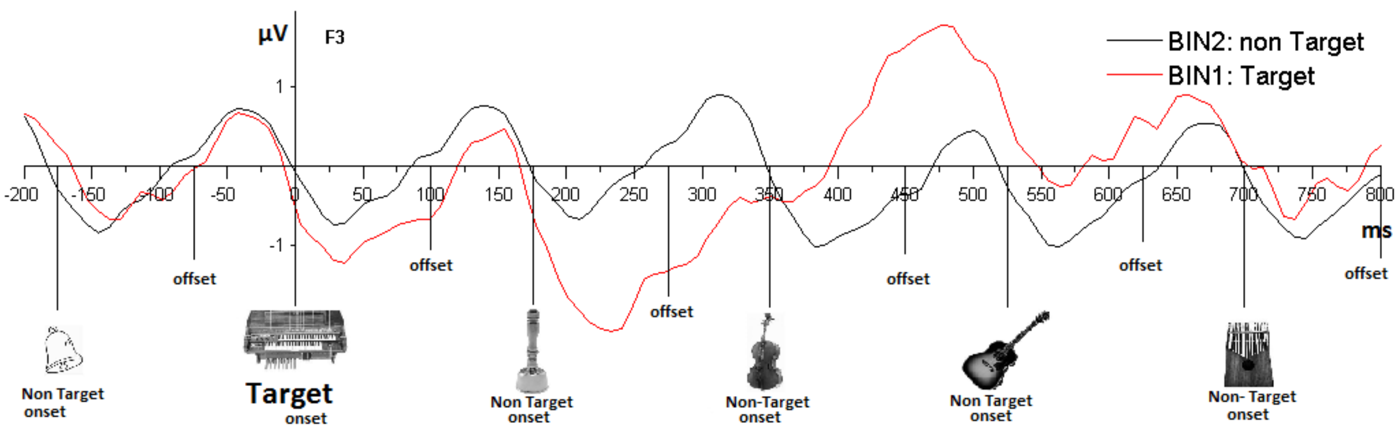


Figure 3

Stimuli presentation of a repetition for the TimPiSp175 condition and averaged ERP response.

The averaged ERP response shown is measured in the F3 channel of all users for the TimPiSp175 condition. The red line corresponds to the target epochs and the black line corresponds to the non-target epochs. The ERP responses follow the periodicity of the stimuli presentation.

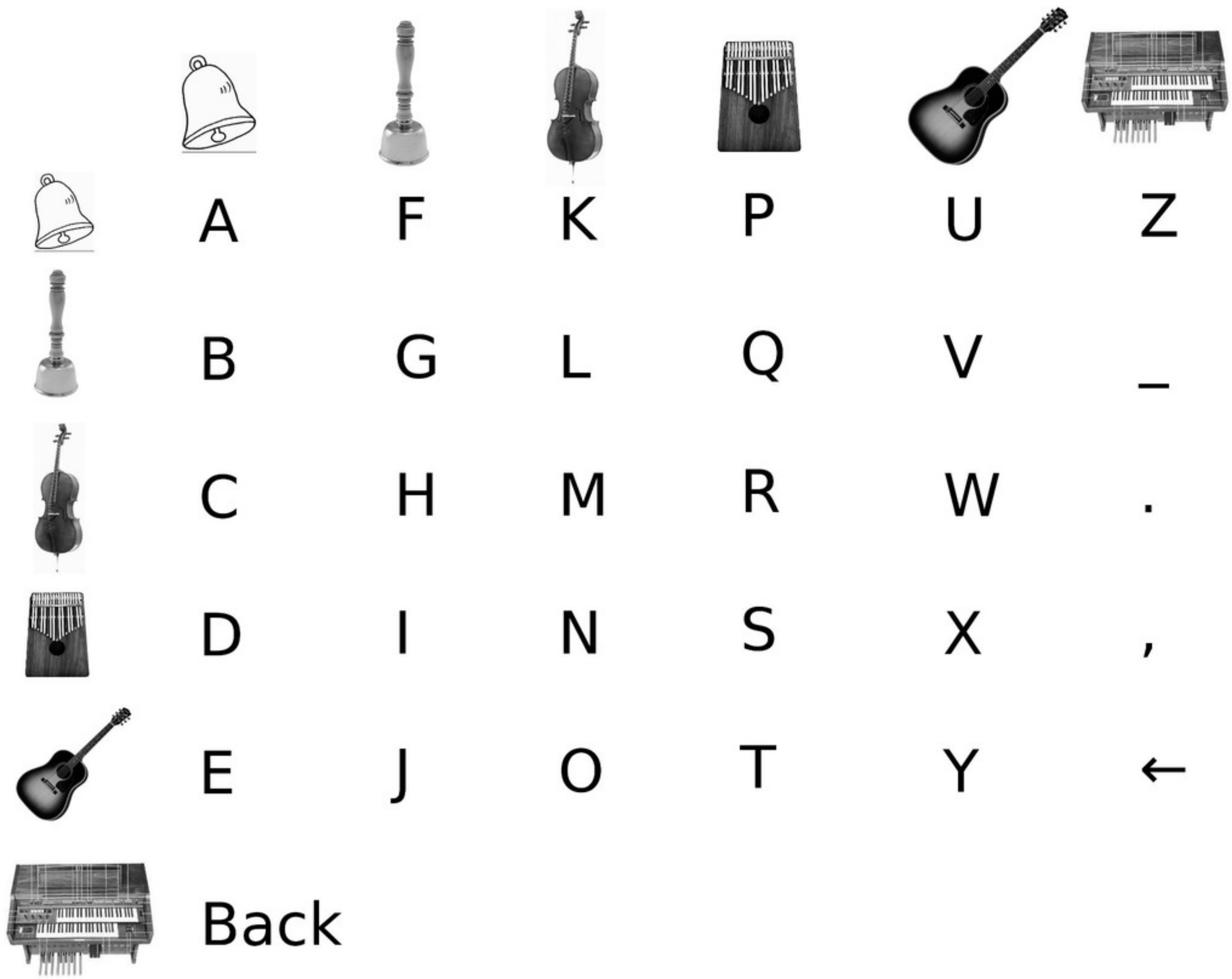


PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 4

Mapping of stimuli into letters

For selecting a particular letter, first the column containing the letter is to be selected (by attending to the corresponding stimulus) and then the row containing the letter is to be selected.

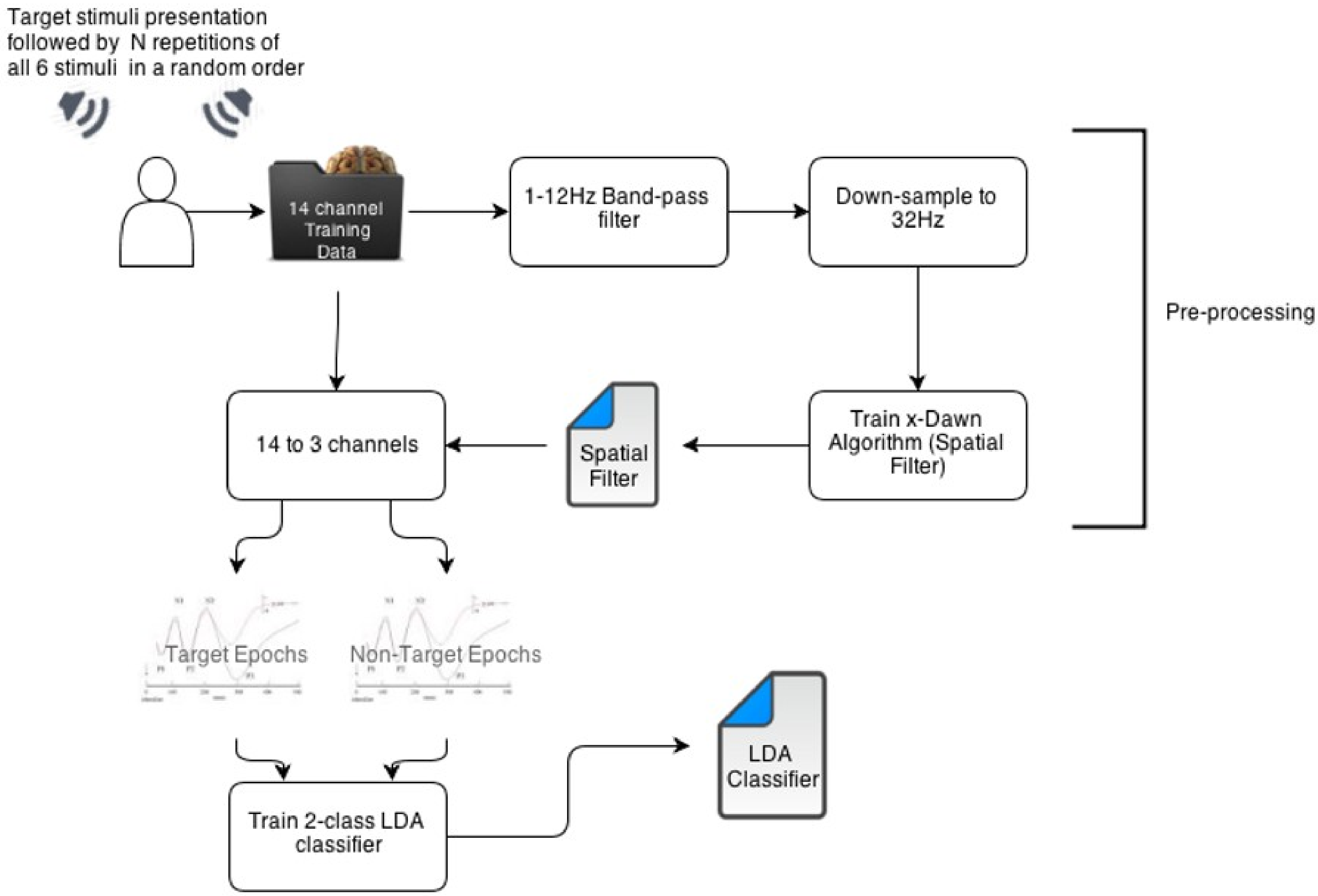


PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 5

Acquiring a Spatial filter and a two class LDA Classifier.

After band-pass filtering (1-12Hz) and down-sampling from 128 to 32Hz, a xDawn algorithm is used to obtain a 14 to 3 channels spatial filter. For each sub-trial a 250 to 750ms after stimulus presentation epoch was created in order to obtain a 48-features vector. The training data consisted of 900 sub-trials (150 target) in the 300ms condition and 1500 sub-trials (250 target) in the 175 and 150ms conditions. Using these data a 2 class LDA classifier was trained to discriminate target from non-target epochs.



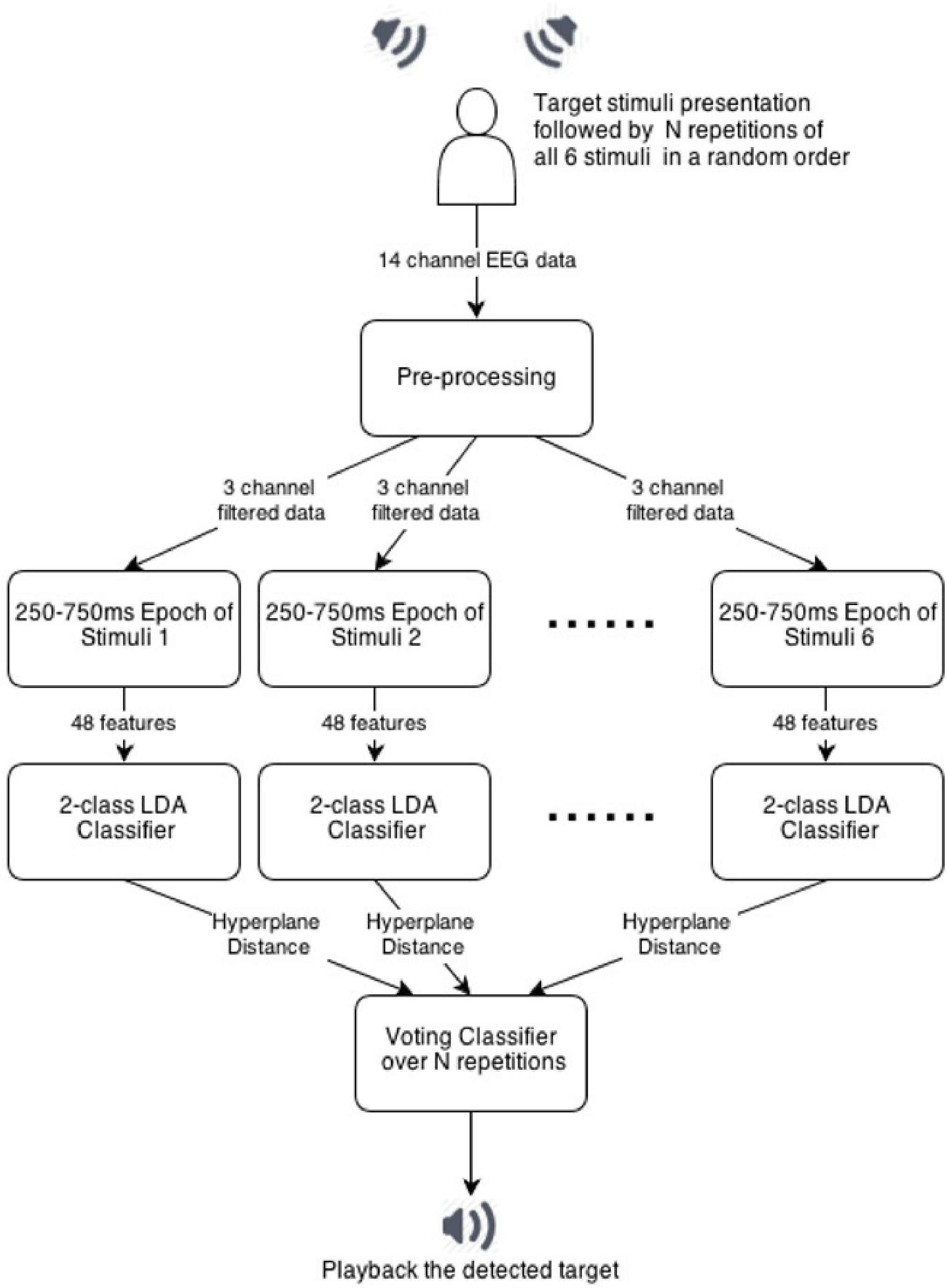
PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 6

Voting Classifier

Similarly to the training session, the online session consists of 10 trials. For every sub-trial, the obtained LDA classifier outputs a hyper-plane distance value. At the end of each trial, a Voting Classifier outputs as target the stimulus that has the minimum sum of Hyper-plane distances over the N number of sub-trials (where N is 15 for the 300ms condition and 25 for the 175 and 150ms conditions).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013



PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 7

On-line performance and ITR for all number of repetitions

(a,b) Averaged on-line performance and ITR of all subjects for the 175 and 150ms conditions for different number of repetitions. (c,d) Averaged on-line performance and ITR of all subjects for the 300ms conditions for different number of repetitions.

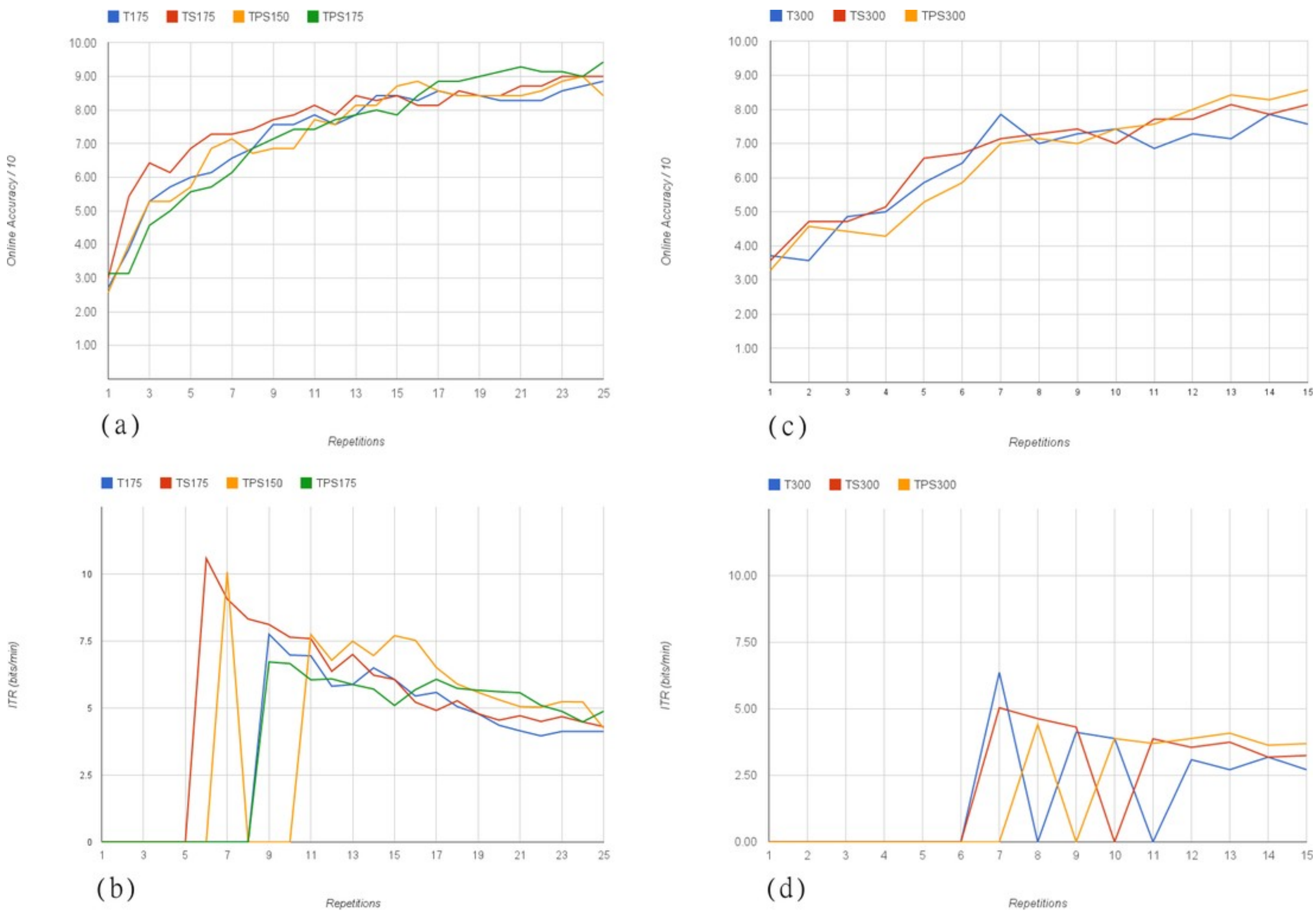
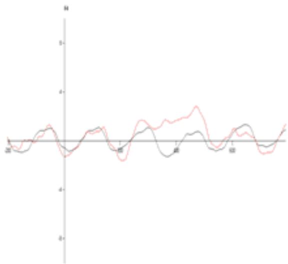
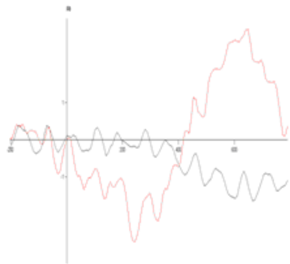


Figure 8

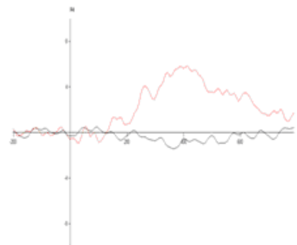
175ms ISI Gran Average



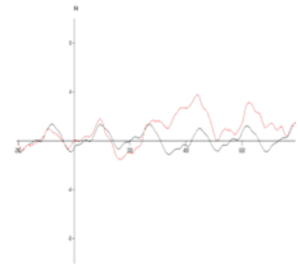
m46 F4



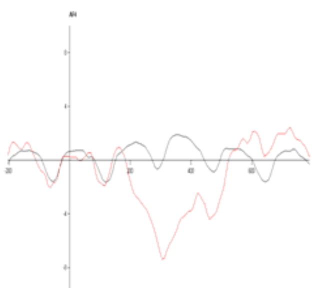
m30 F8



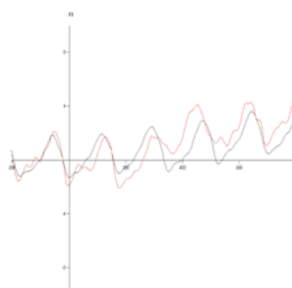
m36 F4



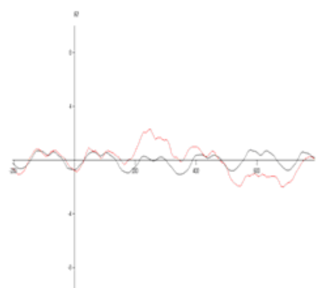
f28 F4



m28 AF4



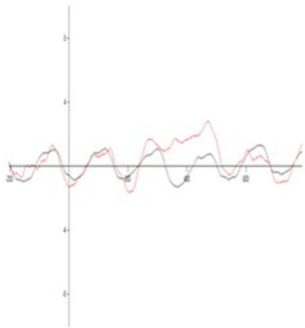
f58 F3



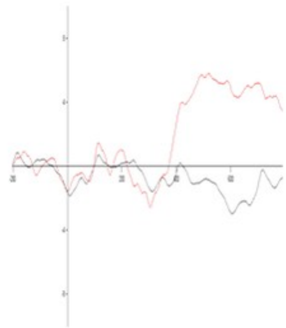
F57 F7

Figure 9

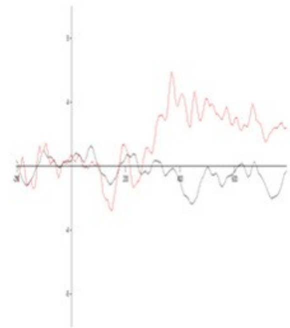
300 ms ISI Gran Average



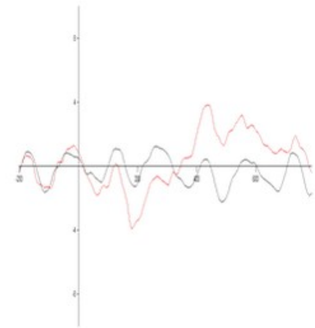
m46 F4



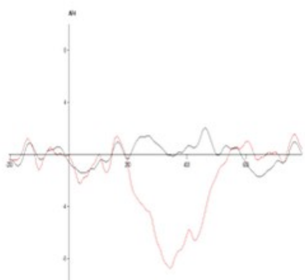
m30 F8



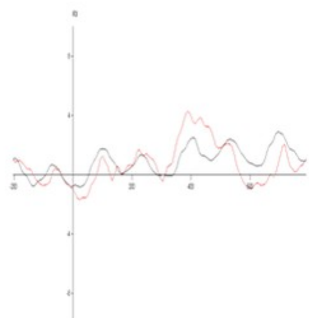
m36 F4



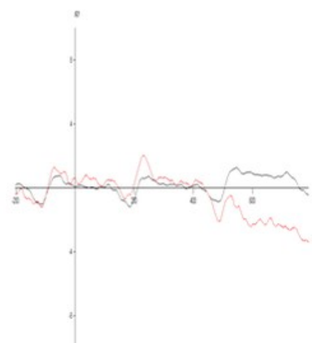
f28 F4



m28 AF4



f58 F3

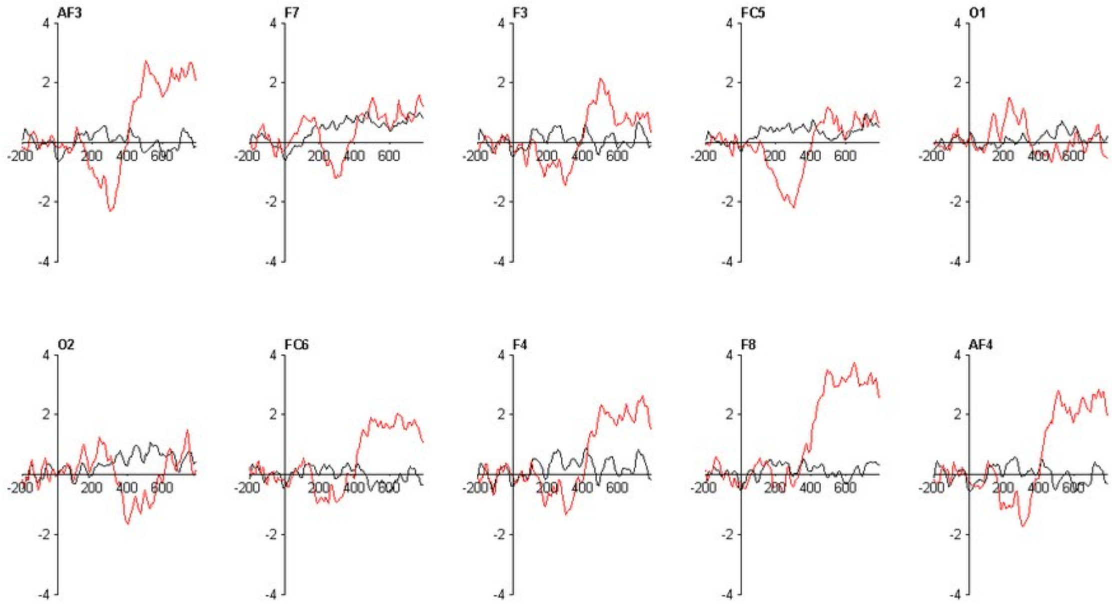


F57 F7

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 10

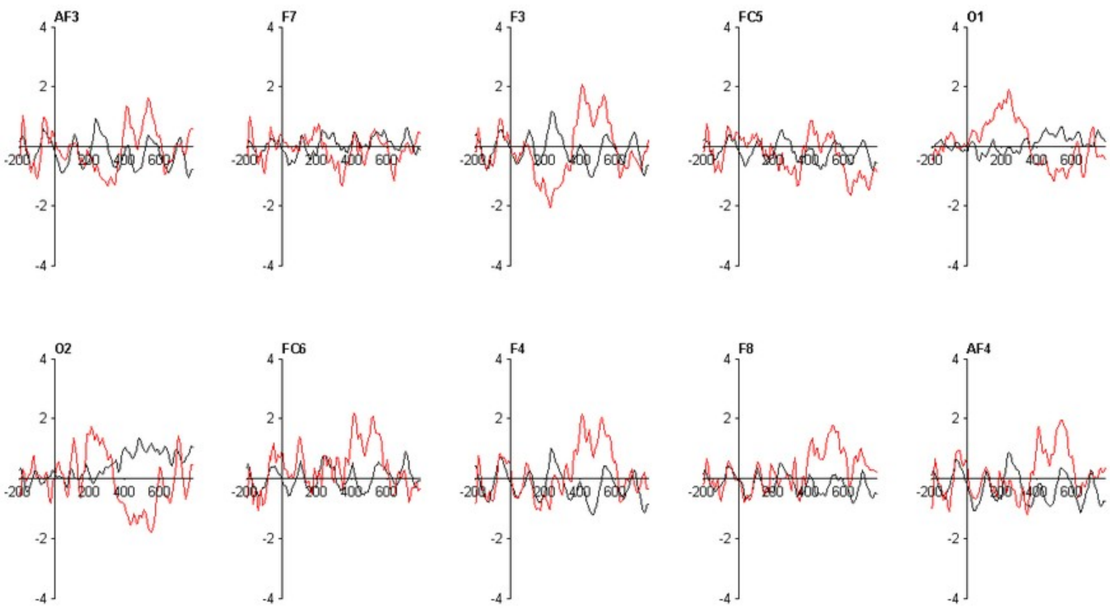
300msTPS condition all subjects 10 electrodes average



PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 11

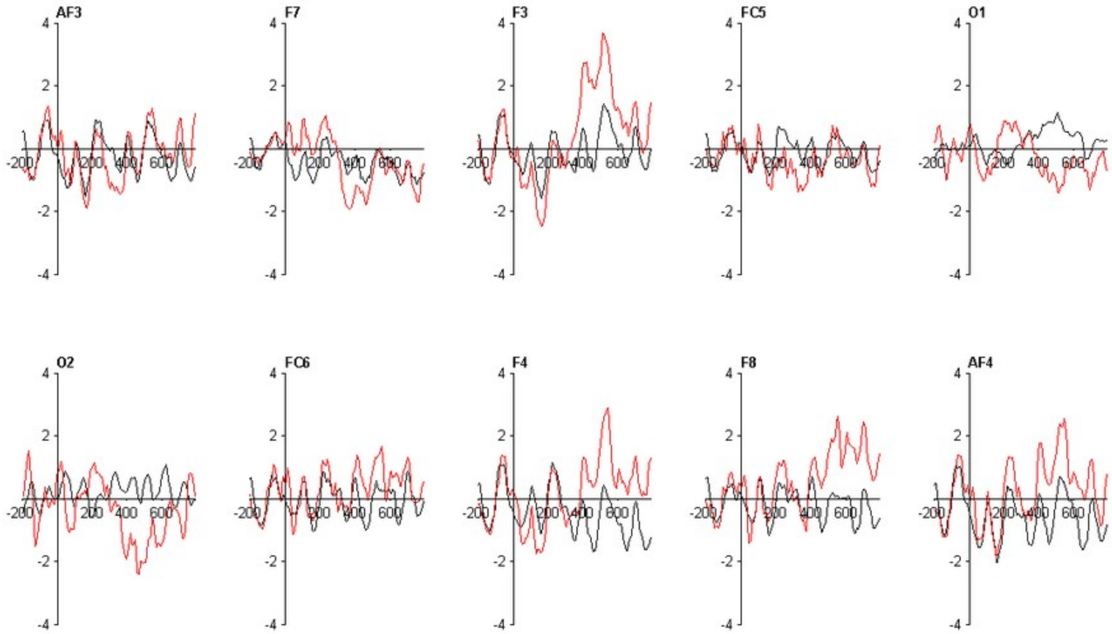
300msTS condition all subjects 10 electrodes average



PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Figure 12

300msT condition all subjects 10 electrodes average



PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Table 1(on next page)

Cue properties in the different conditions

Timb		TimSp			TimPiSp		
Pitch (Hz)	**Stimuli**	**Pitch**	**Stimuli**	**Spatial**	**Pitch**	**Stimuli**	**Spatial**
130.81	Bell	130.81	Bell	-45	23.123	Bell	-45
130.81	Snare Drum	130.81	Snare Drum	-27	51.91	Cello	-27
130.81	Hi Hat	130.81	Hi Hat	-9	116.541	Organ	-9
130.81	Guitar	130.81	Guitar	9	261.626	Guitar	9
130.81	Kalimba	130.81	Kalimba	27	587.330	Kalimba	27
130.81	Claps	130.81	Claps	45	1318.51	Bali bell	45

Table 1: Cue properties in the different conditions.

Table 2(on next page)

Results for Timbre Pitch Spatial (TimPiSp) modality

For each condition and each user is given: (i) the Selection Accuracy and in parenthesis the Number of Repetitions Required, (ii) the Maximum ITR achieved and in parenthesis the Number of Repetitions that maximize it, under the constraint that at least a 70% of accuracy is achieved, (iii) the Amplitude in μV and (iv) the Latency in ms of the P300 peak in the (v) given position and finally (vi) the percentage of rejected epochs during the off-line analysis.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

	300 ms ISI						175 ms ISI						150 ms ISI					
Subject	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)
M30	100 (7)	23,3 (3)	7,28	479	F8	4,2	100 (12)	15,9 (5)	5,24	639	F8	3,8	100 (11)	13.32 (7)	3.55	427	F8	1.2
M46	80 (13)	3,59 (13)	0,13	659	F4	0,2	90 (21)	11,5 (5)	2,36	478	F4	2,1	100 (23)	7.85 (16)	3.64	484	F4	0.7
M36	70 (8)	4,2 (8)	6,95	458	F4	0,9	100 (14)	13,4 (8)	6,99	505	F4	1,4	90 (10)	15.54 (6)	6.66	378	F4	41.5
M28	100 (15)	5,83 (8)	0,079	613	AF4	2,9	100 (25)	5,75 (10)	2,76	597	AF4	2,9	90 (24)	15.54 (6)	4.02	559	AF4	4.2
F28	100 (15)	8,97 (7)	3,53	433	F4	2,7	100 (17)	13,4 (8)	3,14	615	F4	1,9	100 (15)	11.5 (15)	3.07	648	F4	6.3
F58	90 (15)	4,8 (7)	5,18	436	F3	22,6	100 (25)	5,91 (25)	1,79	449	F3	6,9	70 (14)	6.66 (14)	1.42	391	F3	20.4
F57	80 (13)	3,59 (13)	1,25	503	F7	13,4	90 (20)	5 (16)	2,3	248	F7	1,6	100 (23)	33.57 (2)	2.09	267	F7	2.7
Mean	88,5 (12.3)	7,75 (8.4)	3,49	511		6,7	97,1 (19.1)	10,1 (11)	3,51	504		2,94	92.86 (17.1)	14.85 (9.43)	3.49	450.57		11.00

Table 3 (on next page)

Results for Timbre Spatial modality (TimSp)

Fields as in table 1. The ITR is not computed when the limit of 70% accuracy is not reached

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

	300 ms ISI						175 ms ISI					
Subject	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)
m30	100 (4)	33.5 (1)	6.61	519	F8	11.9	100 (9)	28.7 (2)	4.8	606	F8	1.6
m46	100 (7)	16.7 (2)	7.4	426	F4	0.1	90 (20)	6.15 (13)	5.13	466	F4	0.5
m36	60 (13)	-	9.03	369	F4	74.4	100 (9)	15.98 (5)	7.58	388	F4	0.5
m28	100 (6)	14.36 (6)	2.37	602	AF4	2.6	100 (7)	15.98 (5)	2.22	621	AF4	18.3
f28	80 (13)	3.59 (13)	4.83	443	F4	3.3	100 (17)	26.64 (3)	4.1	441	F4	0.7
f58	60 (13)	-	1.83	390	F3	12.7	80 (23)	3.47 (23)	2.77	296	F3	40
f57	90 (15)	4.8 (7)	1.57	245	F7	0.8	70 (10)	5.75 (10)	3	258	F7	2.5
Mean	84.3 (10.1)	10.4 (4.1)	4.8	428		15.1	91.4 (13.6)	14.7 (8.71)	4.2	439		9.16

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Table 4(on next page)

Results for the Timbre modality. Fields as in table 1

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.179v1 | CC-BY 3.0 Open Access | received: 29 Dec 2013, published: 29 Dec 2013

Subject	300 ms ISI						175 ms ISI					
	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70% stable	amplitude (µV)	Latency (ms)	Electrode	Rejected Epochs (%)
m30	100 (5)	20,93 (3)	6,29	589	F8	0,4	100 (24)	9.99 (8)	0,19	649	F8	3,2
m46	70 (12)	2,8 (12)	5,45	546	F4	5,9	70 (14)	4.11 (14)	4,48	421	F4	7,8
m36	60 (14)	-	4,6	403	F4	4,9	100 (5)	39.96 (2)	8,62	341	F4	0
m28	80 (7)	6,66 (7)	2,05	551	AF4	1,5	90 (17)	9.99 (8)	2,59	559	AF4	1,2
f28	90 (14)	6,71 (5)	4,17	415	F4	4,7	90 (14)	19.18 (3)	5,22	382	F4	1,1
f58	80 (7)	6,66 (7)	8,37	392	F3	28,9	90 (17)	9.99 (8)	8,9	641	F3	15,9
f57	80 (10)	4,8 (7)	1,84	316	F7	1,5	80 (20)	4,43 (13)	2,44	409	F7	1,8
Mean	80 (9.9)	6,94 (5.86)	4,68	458,86		6,83	88.57 (15.86)	13.95 (8)	4,63	486,00		4,43

Table 5(on next page)

Auditory Speller Experiment results

Subject	Total Time(minutes)	Condition Used	Online Accuracy (speller)	Expected Accuracy (auditory modality experiment)	Chars/min	ITR (bits/min)
M30	14.24	TimPiSp150	93.33%	100%	0.84	4.37
M46	35	TimPiSp175	76%	100%	0.4	2.59
M36	16.16	TimPiSp150	88.23%	90%	0.74	3.76
M28	25.64	TimPiSp150	81.48%	90%	0.47	3.08
F28	19	TimPiSp150	85%	100%	0.63	3.42
F57	43.7	TimPiSp150	70.65%	100%	0.27	2.17
Average	27.55		82.45 %	96.67%	0.56	3.23

Table 6(on next page)

Summarizing the results of proposed auditory P300 Spelling paradigms

The optimal number of repetitions columns clarifies whether the reported ITR is acquired when computing the optimal number of repetitions to maximize the ITR, when maintaining a selection accuracy of at least 70%

	Discriminating Cues	Optimal Number of Repetitions	Number of Subjects	ISI (ms)	Average ITR (bits/min)
Schreuder et al, 2011	Pitch, Spatial	Yes	16	175	5.26
Current Study	Timbre, Pitch, Spatial	No	7	150, 175	3.23
Käthner et al, 2012	Pitch, Spatial	No	20	420	2.76
Klobassa et al, 2009	Timbre, Pitch	No	5	500	1.86
Furdea et Al, 2009	Speech	No	13	625	1.54