

# Initial results of a high-speed spatial auditory BCI

E.M. Schreuder<sup>a,b</sup>, M. Tangermann<sup>a</sup>, B. Blankertz<sup>a,c</sup>

<sup>a</sup>*Berlin Institute of Technology, Machine Learning Laboratory, Berlin, Germany*

<sup>b</sup>*Neuroscience Campus, VU University, Amsterdam, the Netherlands*

<sup>c</sup>*Fraunhofer FIRST IDA, Berlin, Germany*

Correspondence: E.M. Schreuder, TU-Berlin, FR6-9, Franklinstraße 28/29, 10587 Berlin, Germany.

E-mail: martijn@tu-berlin.de, phone +4930 314 24927, fax +4930 314 78622

## Abstract.

Most P300 BCI approaches use the visual modality for stimulation. For use with ALS patients this might not be the preferable choice because of sight deterioration. Moreover, using a modality different from the visual one minimizes interference with possible visual feedback. Therefore, a multi-class brain-computer interface paradigm is proposed that uses spatially distributed, auditive cues. Ten subjects participated in an offline oddball task with the spatial location of the stimuli being a discriminating cue. Different inter-stimulus intervals of 1000 ms, 300 ms and 175 ms were tested. With averaging over multiple classifier outputs, selection scores went over 90% for most conditions; two subjects reached a 100% correct score. Corresponding information transfer rates were high, up to an average optimal score of 20.99 bits/minute for the 175 ms condition (best subject 37.80 bits/minute). We conclude that the proposed paradigm is successful for healthy subjects and shows promising results that may lead to a fast BCI that solely relies on the auditory sense.

*Keywords:* Brain-Computer Interfaces (BCI), P300, auditory, single-trial classification, EEG

---

## 1. Introduction

Brain-computer interfaces (BCI) are a direct connection between the brain and a computer, without using any of the brains natural output pathways [Wolpaw et al. 2002]. Most BCI research is aimed toward developing tools for patients with severe motor disabilities and paralyzes. This group of potential users could particularly benefit from BCI technology, since output pathways that are normally employed by the brain can no longer be used.

So far, the primary choice of interaction modality is vision. However, the patient's inability to direct gaze, adjust focus or perform eye blinks may prove its use in BCI applications to be difficult. Therefore, other modalities are being explored such as audition [Furdea et al. 2009; Hill et al. 2005; Nijboer et al. 2008; Sellers and Donchin 2006] and touch [Cincotti et al. 2007; Müller-Putz et al. 2006] to make BCI vision independent. Moreover, when using such alternative methods for patients with residual vision, the visual modality could be used exclusively for feedback thereby preventing interaction between feedback and stimulation.

The fastest, non-invasive state of the art BCI systems depend on the P300 response. In most humans it is present without training in response to an attended rare event (oddball paradigm). This attended stimulus elicits a positive deflection in the ongoing brain potential, which in general has the largest amplitude over parietal regions [Conroy and Polich 2007]. It has primarily been used in the visual P300 speller [Farwell and Donchin 1988; Lenhardt et al. 2008; Sellers and Donchin 2006], but its amplitude has also been shown to be attention dependent in auditory mode [Furdea et al. 2009; Hill et al. 2005; Kanoh et al. 2008; Nijboer et al. 2008; Sellers and Donchin 2006].

An oddball paradigm requires the stimuli to differ on at least one physical property, be it pitch, loudness or something else. As the location in space is also a physical asset of sound, it could be used as the discriminating property. In fact [Teder-Sälejärvi and Hillyard 1998] used this spatial aspect to present a subject with seven different oddball streams. They did not use the spatial location to separate targets from non-targets but rather to simultaneously present seven separate oddball tasks. An oddball paradigm based purely on spatial location was used in [Rader et al. 2008], but merely as a training for detecting stimuli from different locations in a later task. No behavioral- or neurophysiological data for this condition is reported.

We hypothesize that adding spatial information to the auditory stimuli is an intuitive way to realize a multi-class auditory BCI. The present study was conducted to confirm that classification of the P300 deflection in response to this spatial information is possible with the required accuracy. The setup would be flexible in the number of classes used and could increase the speed of auditory BCI.

## 2. Material and Methods

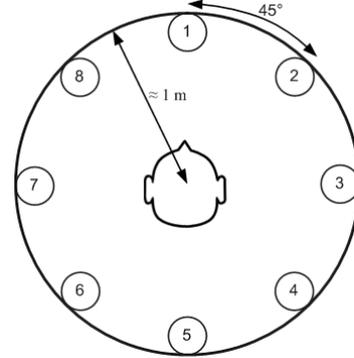
### 2.1 Participants

Ten healthy subjects (six male, mean age 30.5, range 22-55) were tested. Five subjects participated in condition C1000, three in conditions C300 and C175 and two participated in all three conditions (see Table 1 for condition descriptions). Seven subjects reported to have normal hearing, two reported having difficulty with spatial localization of sounds in natural situations and one reported a high-pitched tinnitus in the right ear.

### 2.2 Task Procedure and Experimental Design

Subjects sat in a comfortable chair, facing a screen with a fixation cross. They were surrounded by 8 speakers at ear height. The speakers were spaced evenly with 45° angle between them, at approximately 1 meter distance from the subjects ear (see Fig. 1). Speakers were calibrated to a common stimulus intensity of  $\approx 58$  dB. At the start of a recording session, subjects judged the subjective equality of the loudness from different directions and alter these if necessary. Subjects were asked to minimize eye movements and other muscle contractions during the experiment. The PsychToolbox was used for stimulus presentation [Brainard 1997].

Three experimental conditions were tested, differing in inter stimulus interval (ISI), the amount of speakers used, the stimulus type and the number of repetitions (see Table 1). All conditions consisted of an oddball paradigm with the spatial location of the stimulus being either the only discriminating property (C1000) or a supportive property (C300 and C175). The stimuli in condition C1000 consisted of band-pass filtered white noise and were equal for all directions. In conditions C300 and C175 each direction had a unique tone in addition to the noise, adding a second discriminative property. Prior to every trial, subjects were given a target direction from which they had to count the number of stimulus occurrences. Presentation order was pseudo random with the restriction that all directions were stimulated in one block before continuing to the next. Between two stimuli from the same direction there were always two other stimuli to prevent target overlap.



**Figure 1.** Experimental setup. Subjects are surrounded by 8 speakers. Numbered circles indicate the different speaker locations.

**Table 1.** Settings for the three experimental conditions. The speaker numbers correspond to those given in Fig. 1. One trial consists of a set amount of sub-trials given in the column "Sub-trials"

Condition	ISI (ms)	Speakers	Stimulus	Length (ms)	Trials	Subtrials
C1000	1000	1-8	noise	75	32	80
C300	300	7,8,1-3	noise + tone	40	50	75
C175	175	7,8,1-3	noise + tone	40	40	75

### 2.3 Preprocessing

EEG was recorded monopolarly using a varying number of electrodes. The signals were filtered between 0.1 and 250 Hz with an analog bandpass before being digitized and stored at 1 kHz. For offline analysis signals were low-pass filtered (below 42Hz for visual inspection and below 10Hz for classification) and subsampled at 100Hz. Data was epoched between 150 ms prior to and 800 ms after stimulus onset, using the first 150 ms as baseline. We refer to a single epoch as sub-trial. All sub-trials belonging to one target presentation are collectively referred to as a trial. Sub-trials with a voltage deflection greater than 70  $\mu$ Volt after linear detrending were rejected as artifacts.

### 2.4 Analyses

All  $r^2$  values reported and used are signed  $r^2$  values ( $r^2$  value multiplied by sign of  $r$  value). Twenty channels were selected based on their signed  $r^2$  difference between classes, ten channels with high positive and ten channels with high negative values. Data from these channels was decimated by taking the mean of 5 samples, resulting in 16 post-baseline samples per channel. Each feature vector thus consisted of 320 dimensions. All dimensions were normalized to zero mean and unit variance based on the training set. The normalization vector was stored and used for normalization of the test set.

Classification was done using the Fisher Discriminant (FD) algorithm. Due to the dimensionality of the features (320 dimensions), some form of regularization was advisable. We used the shrinkage

proposed in [Ledoit and Wolf 2004] which counterbalances the systematic error in the calculation of the empirical covariance matrix. A ten-fold cross validation was performed.

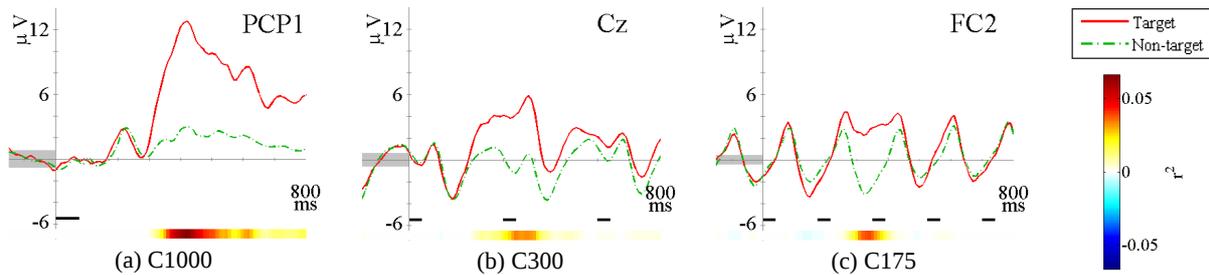
Two types of classification scores can be distinguished: binary classification- and selection accuracy [Furdea et al.]. **(1)** The term (binary) classification accuracy is used for the binary classification of a single sub-trial. It is defined as the percentage of sub-trials that is correctly scored to be a target or non-target. **(2)** The term selection accuracy is used for the percentage of trials in which the target direction is correctly designated.

After cross validation the classifier labels for all sub-trials were used to predict the outcome of the multi-class paradigm, i.e. to estimate the target direction. Taking a block of consecutive sub-trials, one for each direction, the sub-trial with the most negative classifier output was designated the target direction. To increase sensitivity, classifier outcomes for multiple sub-trials for the same direction (within one trial) were averaged. This way, the influence of single sub-trials is decreased. Increased accuracy can be necessary in some applications, where errors are expensive. Because artifacts were rejected classifier labels for some directions were missing. Only the remaining valid sub-trials were used, therefore the averaging for those directions was done over less than the stated number of sub-trials. This is a realistic approach for future online settings, where artifacts may occur at any moment. Various amounts of sub-trials were averaged to evaluate the influence on the outcome.

### 3. Results

#### 3.1 Physiological response

Averaged ERP responses for one subject for all three experimental conditions are given in Fig. 2. For each condition the channel with the highest positive signed  $r^2$  value between 300-650ms is given. For this subject the largest difference is found over the PCP1, Cz and FC2 electrode for condition C1000, C300 and C175 respectively; i.e. with faster stimulus rates the maximal difference between targets and non-targets shifts to more frontal areas. This trend is observed in most subjects.



**Figure 2.** Grand average traces for the different experimental condition for subject VPzq. For each condition, the channel with the highest positive signed  $r^2$  value in the 300-650ms interval is represented. The horizontal color bar indicates the signed  $r^2$  difference between the classes. With the long ISI of 1000 ms (a), a prominent P300 peak is visible for the target condition that peaks around 400 ms. A positive deflection is still observable with the faster stimulus presentations of 300 ms (b) and 175 ms (c), although these are lower in amplitude and  $r^2$  difference. Black bars indicate stimulus presentation.

For condition C1000, all subjects showed an attention dependent positive deflection, albeit with different amplitudes, timing,  $r^2$  values and distributions. For the faster conditions, the positive deflection is superimposed on the rhythmic responses to the ongoing stimuli. Therefore, no typical P300 deflection is observed, but rather a discontinuation of the periodical wave pattern. This is especially true for the fastest condition (see Fig. 2c).

#### 3.3 Classification

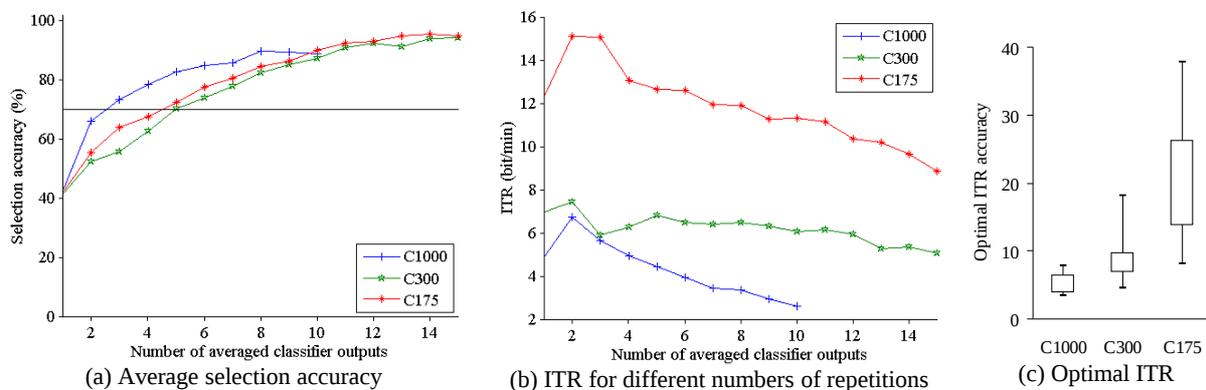
Binary classification scores for all experimental conditions and subjects were 70% or higher. For our unbalanced data set with ratios 1 to 7 and 1 to 4 (target to non-target), a classification score of 87,5% and 80%, respectively, could be obtained by assigning all sub-trials to the non-target class. However, the percentage of correctly classified target and non-target sub-trials was comparable, thus blind labeling of all sub-trials as non-target was not the case.

In the light of BCI, the selection accuracy is of main interest. Chance levels for the selections accuracy are 12,5% and 20% for an eight and five class setup, respectively. When using only a single set of sub-trials all selection accuracies were above change level (range 21,9 – 60). In order to increase accuracy, averaging of classifier outcomes was applied for an increasing number of sub-trials. Selection accuracy quickly rose above 70% for most subjects. The per-condition average selection scores are

presented in Fig. 3a. Generally, the selection scores are highest for condition C1000, although with an increasing number of repetitions the accuracies of all conditions converge. Only one subject did not reach the 70% threshold on any number of repetitions; all subjects in conditions C300 and C175 eventually reached a selection score of 90% or higher.

Closely related to the selection score is the information transfer rate (ITR). Fig. 3b shows the evolution of the average ITR corresponding to the average selection accuracies in Fig. 3a. As average selection accuracies were comparable for all three conditions, the ITR increased with decreasing ISI. ITR for condition C175 is the highest, followed by condition C300. Condition C1000 is initially only slightly worse than condition C300, despite of a more than tripled ISI. It is important to note here that for condition C1000 eight speaker directions were used, in contrast to the five used in the other two conditions. More classes means more information in a single selection.

The optimal number of repetitions varied for all subjects and conditions. Therefore, the lines in Fig. 3b underestimate the optimal outcome. When considering the best outcome for each subject, average best ITR scores were 5.31, 9.47 and 20.99 (best subject 37.80) for conditions C1000, C300 and C175, respectively (see Fig. 3c). Even when only selection accuracies of over 90% are considered the corresponding optimal ITR scores are high: 3.35, 7.05 and 14.41 (best subject 20.60) for conditions C1000, C300 and C175, respectively.



**Figure 3.** Selection accuracy averaged per experimental condition (a) and the corresponding information transfer rate (b). Selections accuracy is generally higher for condition C1000 when compared to the C300 and C175. However, the much faster stimulus presentation in the latter two conditions results in higher average ITR values. When considering the per-subject optimal number or repetitions, the average ITR increases. The distribution of these optimal ITR accuracy for the different conditions is given in (c).

#### 4. Discussion

We discuss here a new experimental paradigm for an auditory BCI. In contrast to most other auditory BCI setups, our setup involves an intuitive multi-class paradigm that can readily vary in the number of classes. From condition C1000 it is evident that the spatial location of a sound can be enough to trigger a P300 response, as it was the only discriminative stimulus property.

The average optimal ITR for condition C175 was 20.99. This is a competitive bit rate when comparing to state of the art auditory BCI systems. [Kanoh et al. 2008] reported an average ITR of around 5 bits/min on their binary BCI, but only when they used all data for training and testing, thereby applying the classifier to data it had already seen. In the binary auditory setup by [Hill et al. 2005], an ITR of 4-7 bits/minute is reported. Our system owes its high ITR to its genuinely multi-class nature. Another multi-class, auditory BCI is reported in [Furdea et al. 2009]. They used spoken numbers as the stimulus for eliciting an ERP. For their 5x6 and 5x5 spelling matrices an average classification accuracy of respectively 40.65% and 60.36% was reported. These are essentially 30- and 25 class BCIs, allowing for high ITRs even when more time is needed for each selection. However, a fair comparison can not be made, as they only report the classification accuracy after using 20 stimulations, which took over 3 minutes to achieve.

The average ITR for condition C175 went down to 14.41 when only 90% correct selection accuracies were considered (best subject 20.60). Although this is a drop in ITR of about 32%, it is a score that is still competitive with other auditory BCI systems and has a much higher accuracy barrier. This high accuracy and corresponding ITR encourage the further development of this paradigm. Several subjects reported that the shorter ISI actually improves their ability to focus on the task.

Visual P300 BCI systems are known for their fast operation and corresponding high ITR. For instance, in a recent online visual speller study [Lenhardt et al. 2008] reported average ITR values of 32.15 bits/minute. For this they used four repetitions with an average classification score of over 80%.

Maximum ITR for a single subject was as high as 92.32 bits/minute using two repetitions. It can be assumed that the reported ITR will go up, when the optimal number of repetitions is determined for each subject individually. Already in the original application of the visual spelling system in 1988 [Farwell and Donchin 1988], ITR values of 12.0 bits/minute (or 10.68 according to [Wolpaw et al. 2000]) were reported. In a direct comparison auditory BCI systems lag behind in their performance. However, the setup proposed in this work takes a step in closing the gap between visual and auditory performance.

## Acknowledgements

This work was partly supported by the European ICT Programme Project FP7-224631 and 216886 and by grants of the DFG (MU 987/3-1) and BMBF (FKZ 01IB001A, 01GQ0850). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

## References

- Brainard DH. The Psychophysics Toolbox. *Spatial Vision*, 10(4):433-436, 1997.
- Cincotti F, Kauhanen L, Aloise F, Palomäki T, Caporusso N, Jylänki P, Mattia D, Babiloni F, Vanacker G, Nuttin M and others. Vibrotactile Feedback for Brain-Computer Interface Operation. *Computational Intelligence and Neuroscience*, Volume 2007:12 pages, 2007.
- Conroy MA, Polich J. Normative Variation of P3a and P3b from a Large Sample. *Journal of Psychophysiology*, 21(1):22-32, 2007.
- Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510-523, 1988.
- Furdea A, Halder S, Krusienski DJ, Bross D, Nijboer F, Birbaumer N, Kübler A. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46, 2009.
- Hill NJ, Lal TN, Bierig K, Birbaumer N, Schölkopf B. An Auditory Paradigm for Brain-Computer Interfaces. *Advances in Neural Information Processing Systems*, 17:569-76, 2005.
- Kanoh Si, Miyamoto K-i, Yoshinobu T. 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*:642-645, 2008.
- Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365-411, 2004.
- Lenhardt A, Kaper M, Ritter HJ. An adaptive P300-based online brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(2):121-130, 2008.
- Müller-Putz GR, Scherer R, Neuper C, Pfurtscheller G. Steady-state somato-sensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(1):30-37, 2006.
- Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland DJ, Birbaumer N, Kübler A. An auditory brain-computer interface (BCI). *Journal of Neuroscience Methods*, 167(1):43-50, 2008.
- Rader SK, Holmes JL, Golob EJ. Auditory event-related potentials during a spatial working memory task. *Clinical Neurophysiology*, 119(5):1176-1189, 2008.
- Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by ALS patients. *Clinical Neurophysiology*, 117(3):538-548, 2006.
- Teder-Sälejärvi WA, Hillyard SA. The gradient of spatial auditory attention in free field: an event-related potential study. *Perception and Psychophysics*, 60(7):1228-1242, 1998.
- Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM. Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164-173, 2000.
- Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767-791, 2002.