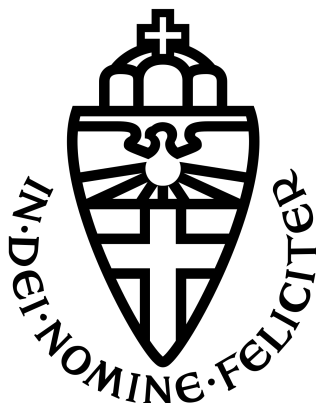


# BACHELOR'S THESIS IN ARTIFICIAL INTELLIGENCE



RADBOUD UNIVERSITY NIJMEGEN

ARTIFICIAL INTELLIGENCE

FACULTY OF SOCIAL SCIENCES

---

## Combining Auditory and Visual Stimuli to Improve the Performance of Covert Attention-based BCIs

---

**Author:**

M.G.M. Driessen  
(Max Driessen)

*Student number: s4789628*

**Supervisor:**

C.S. Verbaarschot  
(Ceci Verbaarschot)

*Donders Centre for Cognition,  
Donders Institute for Brain,  
Cognition and Behavior,  
Radboud University Nijmegen*

July 7, 2019

# Combining Auditory and Visual Stimuli to Improve the Performance of Covert Attention-based BCIs

## Author:

M.G.M. Driessen (Max Driessen)  
Student Number: s4789628

## Supervisor:

C.S. Verbaarschot (Ceci Verbaarschot)  
Donders Centre for Cognition, Donders  
Institute for Brain, Cognition and Behavior,  
Radboud University Nijmegen

## Abstract

Brain-Computer Interface (BCI) spellers allow their users to communicate with the outside world without moving or speaking. The accuracy of well-known BCI spellers such as the matrix speller and the Hex-o-Spell interface has been shown rely on eye gaze (overt attention): if directing gaze to attended options is prohibited, their accuracy drops significantly. In this paper, we perform a pilot experiment to test whether the performance of a visual covert attention-based BCI (in which eye movements are not allowed) can be improved by adding auditory stimuli. We tested two conditions (visual-only vs. visual and auditory stimuli) for a P300-based variant of this type of BCI. For the 4 subjects we tested, adding auditory stimuli did not have a large effect on classifier performance.

## 1 Introduction

Brain-Computer Interfaces (BCIs) are interfaces that allow users to communicate things about their mental state to the outside world, without using their peripheral nervous system (Van Gerven et al., 2009). With a BCI, users can for instance control a prosthetic hand via imagined movement of their left or right hand (Guger et al., 1999). A BCI consists of several components. A measurement device, such as an electroencephalography (EEG) cap, measures some property of the user’s brain activity; in the case of EEG, electrical signals as recorded by (noninvasive) electrodes placed on the scalp. The measured brain signal is then processed by a computer: relevant features are extracted and a classifier is used to predict something about the mental state of the user (e.g. a user’s intention to move his or her hand). The result of this prediction is then fed back to the user (e.g. by moving the prosthetic hand), closing the so-called “Brain-Computer Interface cycle” (Van Gerven et al., 2009), which is shown in Figure 1.

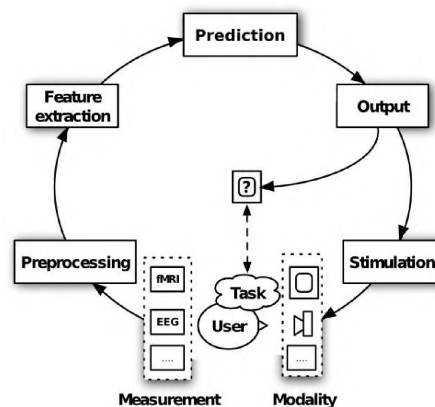


Figure 1: *The Brain-Computer Interface cycle (Van Gerven et al., 2009).*

A common application of BCIs is that of a BCI speller: an interface which allows its users to type words, to communicate with the outside world without speaking or moving. The main purpose of BCI spellers is to restore communication for people who are unable to communicate normally. There are several different types of BCI spellers, a well-known example being the *matrix speller*, introduced by Farwell and Donchin (1988); see Figure 2. In this EEG-based BCI speller, a user is presented with a grid (“matrix”) of symbols. The rows and columns of this grid “flash” in a random pattern (a flash being a short increase in brightness), and the user has to silently count how often the symbol they want to type is intensified (flashed). The BCI attempts to retrieve which letter the user was attending to, and this letter is “typed” at the top of the interface.

MESSAGE					
BRAIN					
Choose one letter or command					
A	G	M	S	Y	*
B	H	N	T	Z	*
C	I	O	U	*	TALK
D	J	P	V	FLN	SPAC
E	K	Q	W	*	BKSP
F	L	R	X	SPL	QUIT

Figure 2: *The visual matrix speller (Farwell & Donchin, 1988). The word that the user typed (“BRAIN”) is shown at the top of the interface.*

The matrix speller relies on a particular component of the *Event-Related Potential* (ERP), that is, the signal that is measured after each “event” (each flash of a row/column). This component, the *P300 component*, is a positive peak in the EEG that occurs approximately 300 milliseconds after an infrequent “oddball” stimulus that someone is attending to (Polich, 2007); see Figure 3.

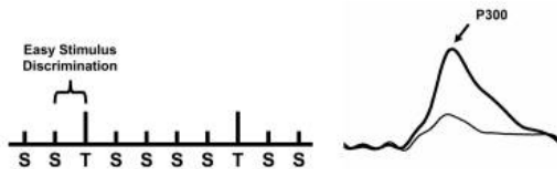


Figure 3: *The P300 component (Polich, 2007). The P300 is elicited by rare “target” events (T) in a stream of “standard” events (S); in the matrix speller, target events are flashes of the row/column that contains the attended symbol.*

In the context of the matrix speller, a P300 component is produced when the attended row or column (containing the letter the user is trying to type) is intensified, which is a relatively rare event compared to how often a non-attended row is flashed. The BCI can detect the presence of the P300 component, and can find out which intensifications elicited it; using this information, it can find which row and column the user was attending to, to retrieve the letter that should be typed.

The main problem of the matrix speller is that it relies heavily on eye gaze (so-called “overt attention”): due to several peculiarities

of the interface such as there being a large number of small symbols, a user needs to look directly at the symbol that he or she is trying to type in order for the system to be able to accurately retrieve the attended symbol (Brunner et al., 2010; Treder & Blankertz, 2010). However, people who are impaired in or completely lacking the ability to communicate with the outside world (e.g. people with locked-in syndrome or Amyotrophic Lateral Sclerosis (ALS)), who are the target population for BCI spellers, often also have difficulty moving their eyes, or doing so for prolonged periods of time (Brunner et al., 2010).

A different type of BCI speller that could help alleviate this problem is the *Hex-o-Spell interface*, introduced by Blankertz et al. (2006). In this BCI speller, symbols are grouped together in six groups of five symbols, which are presented in a circle around a central point. In the original version of the Hex-o-Spell interface, shown in Figure 4a, users can control an arrow in the center of the interface by using imagined movements. Once the arrow touches one of the groups, the symbols contained within this group move to separate locations. The arrow resets so that users can proceed to select one of these symbols, which is then typed.

Several variations of the Hex-o-Spell interface exists, such as one that uses several components of the ERP including the P300 component (Treder & Blankertz, 2010); this variant is shown in Figure 4b. In this ERP-based variation, there is no arrow in the center, but instead the options (groups or separate symbols) are intensified in random order by increasing their size, just like how the rows and columns of the matrix speller are intensified by increasing their brightness. To select an option, users silently count how often this option is intensified; retrieving the attended option is done in the same manner as retrieving a row/column combination in the matrix speller.

The fact that the Hex-o-Spell interface has fewer, larger symbols that are arranged in a circle around a central point could make it more usable for people who have trouble controlling overt attention (eye gaze). In an experiment by Treder and Blankertz (2010), the ERP-based variant of the Hex-o-Spell interface was compared to the matrix speller in both “overt” and “covert” conditions (i.e., in situations where participants were or were not allowed to direct their gaze to the targets). The results are shown in Figure 5.

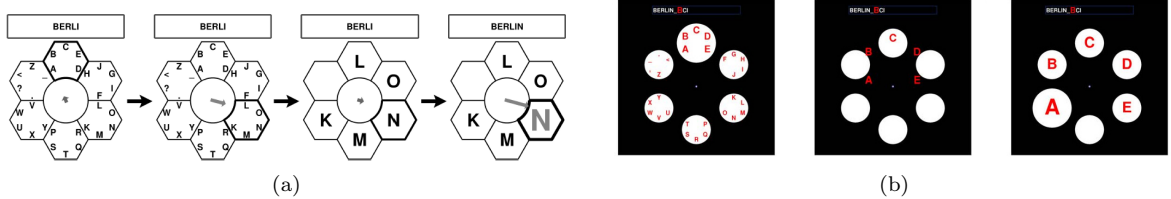


Figure 4: *The Hex-o-Spell interface. (a) The original interface (Blankertz et al., 2006). In this example trial, the group containing symbols “k” through “o” is selected; these symbols are expanded to their own locations, after which the “n” is selected. (b) The ERP-based Hex-o-Spell interface (Treder & Blankertz, 2010). The first image shows an intensification of the top group, which is selected, after which each symbol is moved to its own location. The final image shows an intensification of the “a” symbol.*

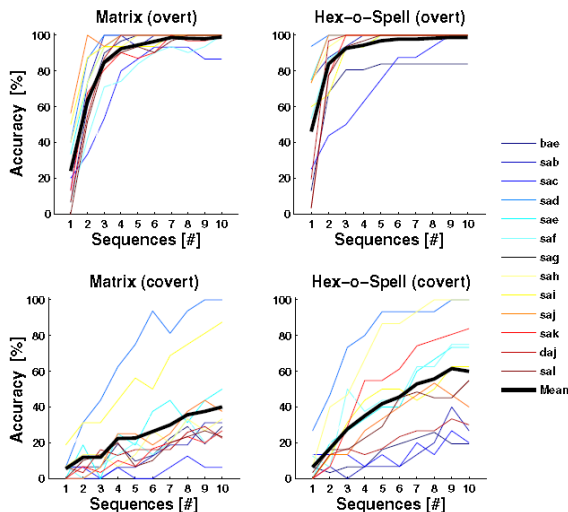


Figure 5: *Classifier accuracy for the matrix and Hex-o-Spell interfaces in the “overt” and “covert” attention conditions (Treder & Blankertz, 2010).*

As Figure 5 shows, both the matrix speller and the Hex-o-Spell interface perform worse in the “covert” condition than they do in the “overt” condition. The Hex-o-Spell interface does outperform the matrix speller in the covert condition by about 20 percentage points after 10 training sequences, but its classifier accuracy in the covert condition (approximately 60%) was found to still be too low for it to be suitable for clinical practice (Treder & Blankertz, 2010).

In an experiment conducted by Belitski et al. (2011), the effects of multi-modal stimuli were investigated by using a variation of the matrix speller. In this variation, only the columns of the matrix are intensified; after a while, the matrix is flipped so that the rows become columns (see Figure 6a).

By combining this version of the matrix speller with six speakers to add localized auditory stimuli to the intensifications of each column, intensifications can be either visual, audiovisual or audio-only. They found that the average ERP response to a target stimulus was highest in the audiovisual condition (i.e. when intensifications consisted of a visual flash as well as an auditory stimulus), as can be seen in Figure 6b. This increase in stimulus response for multi-modal (i.e. audiovisual) stimuli then lead to an increase in classifier performance as well.

In the experiment by Belitski et al. (2011) described above, participants were allowed to direct overt attention (eye gaze) towards the targets. However, the results they found could perhaps be used to increase the accuracy of covert attention-based BCI spellers (such as the ERP-based Hex-o-Spell interface where eye movements are prohibited) as well. In this thesis, we present a pilot study that investigates this possibility: we set up a basic visual covert attention-based BCI, and test whether adding auditory stimuli to this BCI improves its accuracy.

We hypothesize that the results found by Belitski et al. (Belitski et al., 2011) should also hold for BCIs where eye movements are not allowed. Hence, we expect to see an increase in stimulus response, and therefore an increase in classifier performance, if auditory stimuli are used in addition to the visual stimuli of such a BCI. If our hypothesis is correct, this result could be extended to more substantial covert attention-based BCIs such as the ERP-based Hex-o-Spell interface. This could increase their accuracy and make them more usable by their target population, hence making these sorts of systems more viable for clinical use.

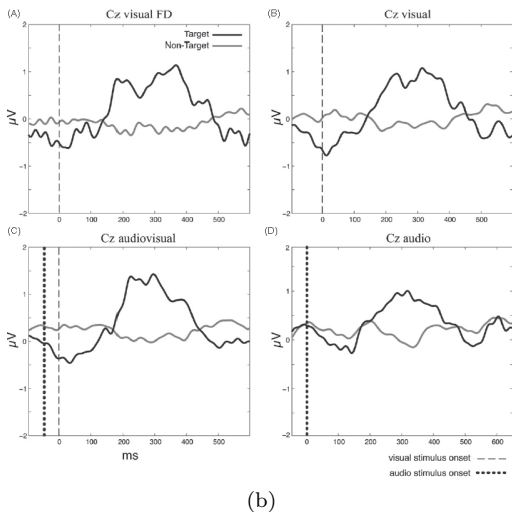
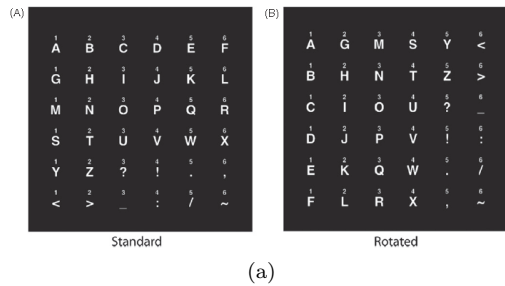


Figure 6: *The matrix speller and results from the experiment by Belitski et al. (2011). (a) The variation of the matrix speller that was used in the experiment. The image on the right shows the “flipped” matrix. (b) The average ERP responses (target vs. non-target (distractor); channel Cz). The top-left plot corresponds to the original matrix speller (Farwell & Donchin, 1988). As can be seen in the three remaining plots, the ERP response in the “audiovisual” condition was more pronounced than those in the “visual” and “audio” conditions.*

## 2 Methods

To test our hypothesis, we conducted a pilot experiment in which we investigated two versions of a covert attention-based BCI. Using this BCI, participants had to answer simple yes/no (i.e. closed) questions, and the BCI gave them online feedback. One version of the BCI only used visual stimuli, while the other combined the same visual stimuli with auditory ones. The results of these two conditions were then compared to see whether adding the auditory stimuli improved classifier performance.

## 2.1 Participants

Four participants (1 male, 3 females, aged 20-21) took part in the pilot study. All participants were Dutch nationals who were proficient in English. None of the participants had hearing problems or color blindness; all had normal or corrected-to-normal vision. All participants were properly informed beforehand, and gave written consent.

## 2.2 Data Acquisition

EEG was recorded using a Biosemi system with 64 AgCl active electrodes, which were placed on the scalp according to the international 10-20 system. Data was sampled at a rate of 256 Hz, and processed using the Buffer BCI toolbox (Buffer BCI, 2019). In addition to the EEG channels, 4 EOG channels were placed (next to each eye, and above and below one eye), as well as 2 mastoid channels.

## 2.3 Interface

For our experiment, we created a BCI that allowed participants to answer simple yes/no questions by counting the number of “target intensifications” of the desired option without directing their gaze to this option. We limited the number of options to 2 to keep the experiment simple, but chose a setup that should make our results extendable to interfaces with more options, such as the ERP-based Hex-o-Spell interface (Treder & Blankertz, 2010). The questions were created specifically to be quickly decidable by participants (e.g. “Does 1 plus 2 equal 3?”), while keeping them engaged. The purpose of the questions was to prime participants to give certain answers, to facilitate retrieval of the attended options in our analysis. The complete list of all questions is shown in Appendix A.

Figure 7 shows a single trial of our experiment (i.e., the process of answering a single question). At the start of a trial, a question is shown for 3 seconds, during which the participant has to read the question and decide on an answer (yes or no). After these three seconds, a red fixation cross appears in the center of the screen, together with two options (“YES” and “NO”). These options are located approximately 12.5 centimeters left and right of the cross, respectively, so that they are in the peripheral part of the participant’s visual field.

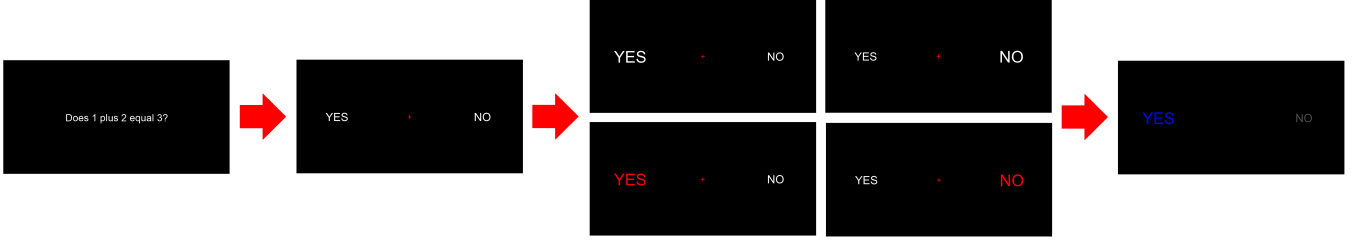


Figure 7: A single trial of our experiment. A question is displayed, after which the options are shown. The options start intensifying in random order, with 2-4 “distractors” (top two images) between every two “targets” (bottom two images). In total, there are 10 target intensifications. Finally, after the intensifications are done, the fixation cross disappears and the classification result is fed back to the participant.

After the options have been on screen for one second, the intensifications start. The options are intensified in random order, similar to the intensifications in the matrix speller (Farwell & Donchin, 1988) and ERP-based Hex-o-Spell interface (Treder & Blankertz, 2010). However, since our interface only has two options, our interface requires two types of intensifications: *distractors* and *targets*. Each trial contains 10 target intensifications, with 3-7 of these being intensification of the “YES” option, and the remaining ones being those of the “NO” option. The precise order and distribution of these target intensifications is randomly determined. The task given to each participant is to count how many target intensifications there were for the option corresponding to the answer to the question; this should be done without directing eye gaze away from the fixation cross.

Between every two target intensifications, as well as before the first and after the last one, between 2 and 4 distractor intensifications take place. These distractors mimic the effects of intensifying a non-attended row/column in the matrix speller (Farwell & Donchin, 1988) or a non-attended group in the Hex-o-Spell interface (Treder & Blankertz, 2010), and serve to make the target intensifications “oddball” stimuli, leading to P300 responses to target intensifications of the attended option (Polich, 2007).

In the part of our experiment where only visual stimuli are used (see Section 2.4), an intensification consists of one of the options rapidly increasing in size (by approximately 40%) for 150 milliseconds, after which it returns to normal. There are 150 milliseconds between the end of one intensification and the start of the next one. During distractor intensifications, the intensified option remains white, whereas during target intensifications, the option turns red.

In the “audiovisual” part of our experiment, intensifications are very similar to the ones in the “visual” part, but in addition to the

visual stimuli, auditory stimuli are played via earphones. During a target intensification, a 500 Hz beep is played in the ear corresponding to the option that is being intensified (left ear for “YES”; right ear for “NO”), for the entire duration of the visual intensification (150 ms). Distractor intensifications are accompanied by a beep as well, but with a frequency of 400 Hz. This was done to make the target intensifications “oddball” stimuli in the auditory modality as well as the visual modality, as was also the case in the experiment by Belitski et al. (2011). The frequencies of the beeps were specifically chosen to be pleasant to listen to for prolonged periods of time.

During the experiment, the BCI gathered data for one second after every target intensification. During the testing phases of both parts of our experiment, a classifier was used to determine which of the two options the participant was attending to (see Section 2.5). One second after the intensifications stopped, the prediction made by the classifier was fed back to the participant by changing the color of the options (as shown in Figure 7). The feedback remained on screen for two seconds, after which the trial ended. During the training phase, no feedback was provided. so trials ended one second after intensifications ended.

## 2.4 Procedure

Before the experiment, each participant was given a document containing an explanation of the experiment; this text can be found in Appendix B. During the experiment, a participant sat in a booth, in front of a computer screen. The distance between participant and screen was approximately 60 centimeters. The experimenter sat outside of this booth in a separate room, and could observe participants via a monitor connected to a video camera.

In our experiment, two conditions were tested: a *visual* condition and an *audiovisual* one, the only difference between the two being that in the visual condition only the visual stimuli were present, whereas in the audiovisual condition, the auditory stimuli described in Section 2.3 were added. The task in both conditions was to answer each question by counting how often the corresponding option (“YES” or “NO”) was intensified, without directing eye gaze away from the fixation cross. Once the experiment started, participants were told which condition they would begin with. Participants 1 and 4 started with the visual condition; 2 and 3 with the audiovisual condition.

Each condition consisted of a *training phase* and a *testing phase*, which both consisted of a random selection of 10 questions such that the correct answer to 5 of these questions was “yes”, and the answer to the other 5 was “no”. Throughout the experiment, a participant never encountered the same question more than once (except for participant 1, due to a coding error, but this did not influence our results). There was a break of 5 seconds after every two questions, and after each phase there was a longer break during which the experimenter checked in on the participant.

During the training phase of both conditions, no feedback was provided; this phase served as a way of gathering data on which a classifier could be trained (see Section 2.5). After the training phase, the testing phase began, during which online feedback was provided to the participants in the manner described in the last paragraph of Section 2.3. This phase was used to test the accuracy of the trained classifier.

Once the test of the first condition was completed, there was a short break, after which the remaining condition was tested. The procedure for the test of this second condition was identical to that of the first part, except for the stimuli that were used. Upon completion of this second part, participants were asked to describe their opinion of the experiment, and had to fill in a questionnaire to submit the answers they tried to give to the questions, to allow us to correct our analysis for any wrong answer that was given.

## 2.5 Classifier

We used a linear classifier to detect the P300 component, which then allowed us to predict which option a participant was attending to. We opted for a linear least squares classifier that uses generalized cross-validation (Scikit-learn, 2019). This type of classifier was chosen for the sake of simplicity and because the decision the classifier has to make is a binary one: for every data point, it has to classify it as belonging to either the attended option (“P300 detected”) or the unattended one (“no P300 detected”).

For each condition of our experiment, we trained the classifier on the data obtained during the training phase, which consisted of 100 data points, each of which contained one second of EEG recordings (one data point for each target intensification; each question had 10 target intensifications and there were 10 questions in each training phase). The trained classifier was then used to determine the feedback that was given to the participant during the testing phase of that particular condition. After every target intensification, one second’s worth of EEG data was given to the classifier, which then returned a prediction of whether the intensification belonged to the attended or the unattended option (average predicted value 0.5, with higher values corresponding to “attended”/“P300 detected” and lower values to “unattended”/“no P300 detected”). The resulting predictions were then averaged per option, to find an average prediction value for “YES” and “NO”. The option with the highest average prediction value was then selected as the attended option and fed back to the participant.

## 2.6 Analysis

To preprocess the obtained data, the data was detrended, and bad channels were removed using outlier detection on the channel dimension. The data was re-referenced to a Common Average Reference (CAR); this reference was used because it is properly implemented in the Buffer BCI toolbox (Buffer BCI, 2019) and has been found not to differ significantly from an ear reference for P300 detection (Krusienski et al., 2008). Next, a bandpass filter (0.1-48 Hz) was applied to filter out electrode drift and line noise. Finally, bad trials were removed by using outlier detection on the trial dimension.

After we had obtained data from all participants, we analyzed the data offline. We checked the results of the questionnaire (see Section 2.4), and investigated the obtained ERPs as well as classifier accuracy.

### 2.6.1 ERPs

To obtain ERP plots, we preprocessed the data as described above. We then combined the training and testing data for each participant-stimulus condition combination, and plotted the ERPs that resulted from averaging this data in a layout resembling that of the placement of the electrodes on the scalp.

Next, we zoomed in on channel Cz, to be able to see in more detail whether adding auditory stimuli resulted in an increase in stimulus response, as we hypothesized in Section 1. This was also done in the experiment by Belitski et al. (2011); see Figure 6. In Section 3.2, we show the ERPs found in our experimental conditions.

Aside from doing this for each participant separately, we also created *grand average* ERP plots using data from all participants.

### 2.6.2 Classifier

The classifier that was used during the experiment itself was erroneously trained on the data obtained from the EOG and mastoid channels in addition to the EEG data. To obtain more realistic results, we therefore had to reconstruct our classifier in an offline simulation where we only used EEG as training data. In this simulation, data was preprocessed in the same manner as was done in the online experiment, the only difference being that the EOG and mastoid channels were left out. The predictions made by the resulting classifier were very similar to those made during the online experiment: only a few predictions were noticeably different.

To retrieve the performance of our classifier, we trained and tested the classifier for every participant-stimulus condition combination, in the same manner as was done in the online experiment. To obtain a better representation of

classifier performance, we also ran 5-fold cross-validation: in each fold, a classifier was trained on 16 out of the 20 (training + testing) questions of the corresponding condition, and tested on the remaining 4. In Section 3.3, we report both how often the classifier gave the correct feedback, and how many individual data points (brain signals resulting from intensifications) it classified correctly.

To test whether the classifiers actually used the P300 response to predict whether data points belonged to the attended option or the unattended option, we trained another classifier on the data of participant 1, using only the 9 channels where we would expect a P300 response: C1, Cz, C2, CP1, CPz, CP2, P1, Pz and P2 (i.e., the channels around the centroparietal midline (Wood et al., 1980)). The resulting classifier was tested in the same way as was done for the regular classifier, by using 5-fold cross-validation.

## 3 Results

### 3.1 Questionnaire

The results of the questionnaire (see Section 2.4) showed that all questions had been answered correctly by all participants. Hence, we could use the correct answer to each question to find which of the two options each participant attended to during the intensifications belonging to that question.

### 3.2 Stimulus Response

#### 3.2.1 Participants

Figure 8 shows the data obtained from participant 1: ERP plots of all (non-bad) channels in a layout resembling the placement of the corresponding electrodes on the scalp. As can be seen in these plots, the channels where we would expect a P300 component to appear (i.e. those around the centroparietal midline (Wood et al., 1980)) do indeed show a P300 component.

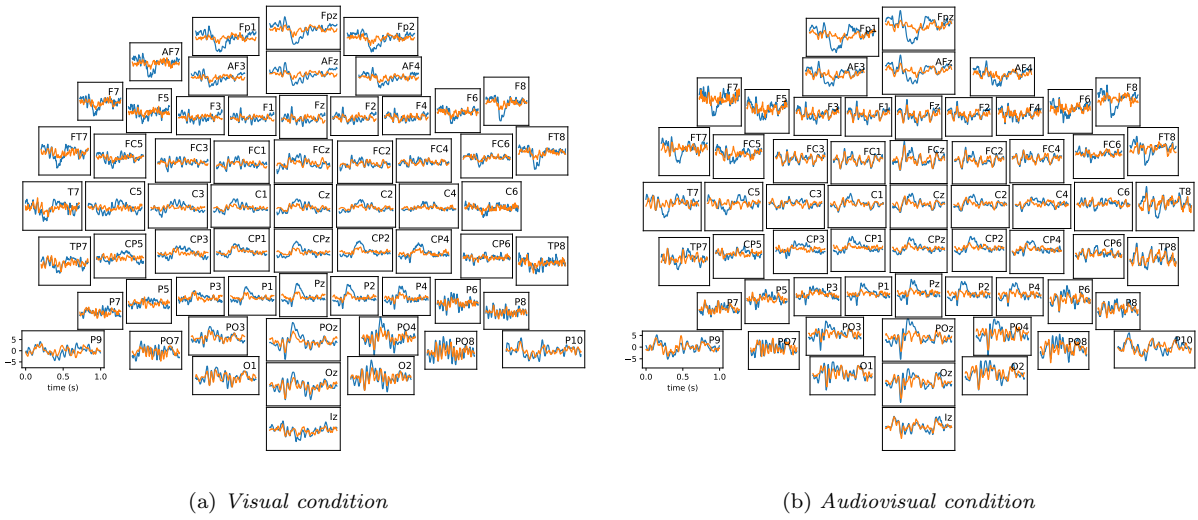


Figure 8: ERPs obtained from participant 1. Blue lines correspond to intensifications of the attended option; orange lines to those of the unattended option. Note that bad channels were left out of these plots. Channels of interest are C1, Cz, C2, CP1, CPz, CP2, P1, Pz and P2.

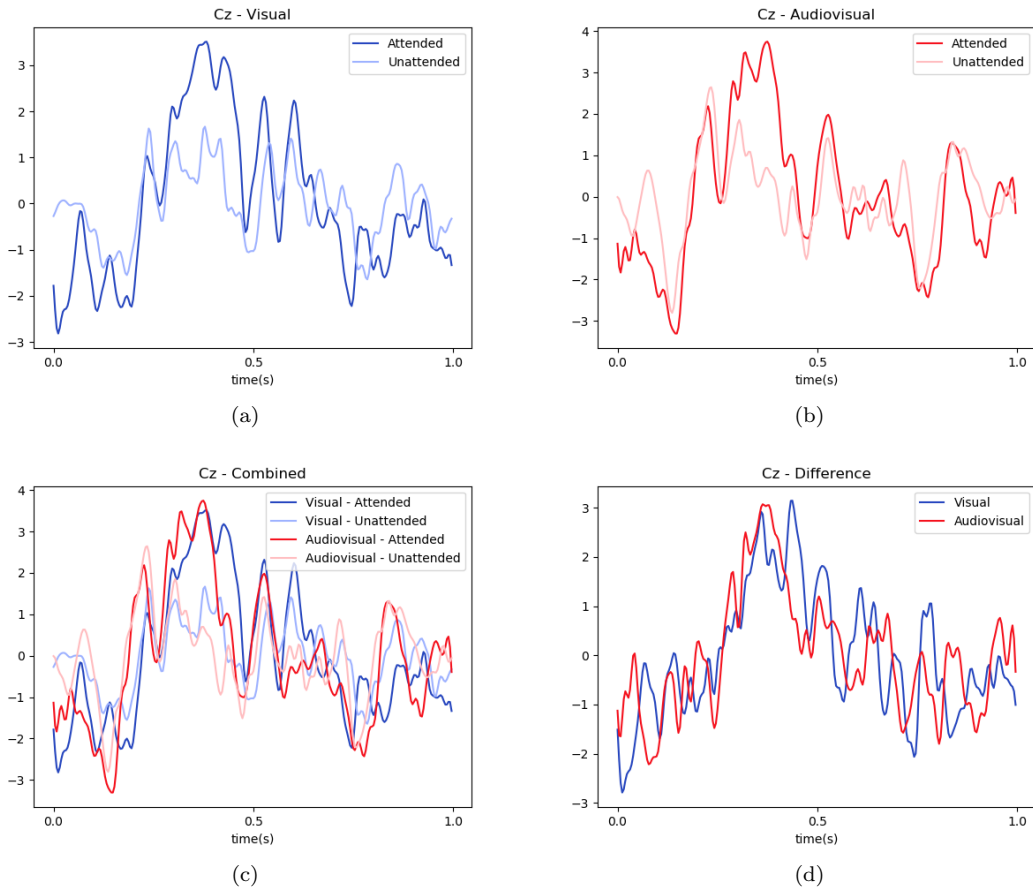


Figure 9: ERPs of channel Cz for participant 1. (a-b) Separate ERP plots (attended vs. unattended option) for each condition. (c) The ERP plots combined in one figure. (d) The difference between the attended and unattended ERP for each condition (for a better comparison).

The plots resulting from the investigation of the data obtained from channel Cz for participant 1 are shown in Figure 9. The ERP responses of the visual and audiovisual conditions were plotted in separate figures, as well as together in a single one. Finally, the difference between the ERPs belonging to the attended and unattended options was computed; the result of this was plotted to allow for a better comparison between the two conditions.

Similar figures were created for the other participants. As they do not add much to the current discussion, these will not be shown here. Instead, they can be found in Appendix C (Figures 13-18).

### 3.2.2 Grand Average

Figure 19 in Appendix C shows the ERP plots of all channels in the computed grand averages. The grand average ERPs found for channel Cz are shown in Figure 10 below, in the same manner as was done for the data obtained from participant 1 in Figure 9.

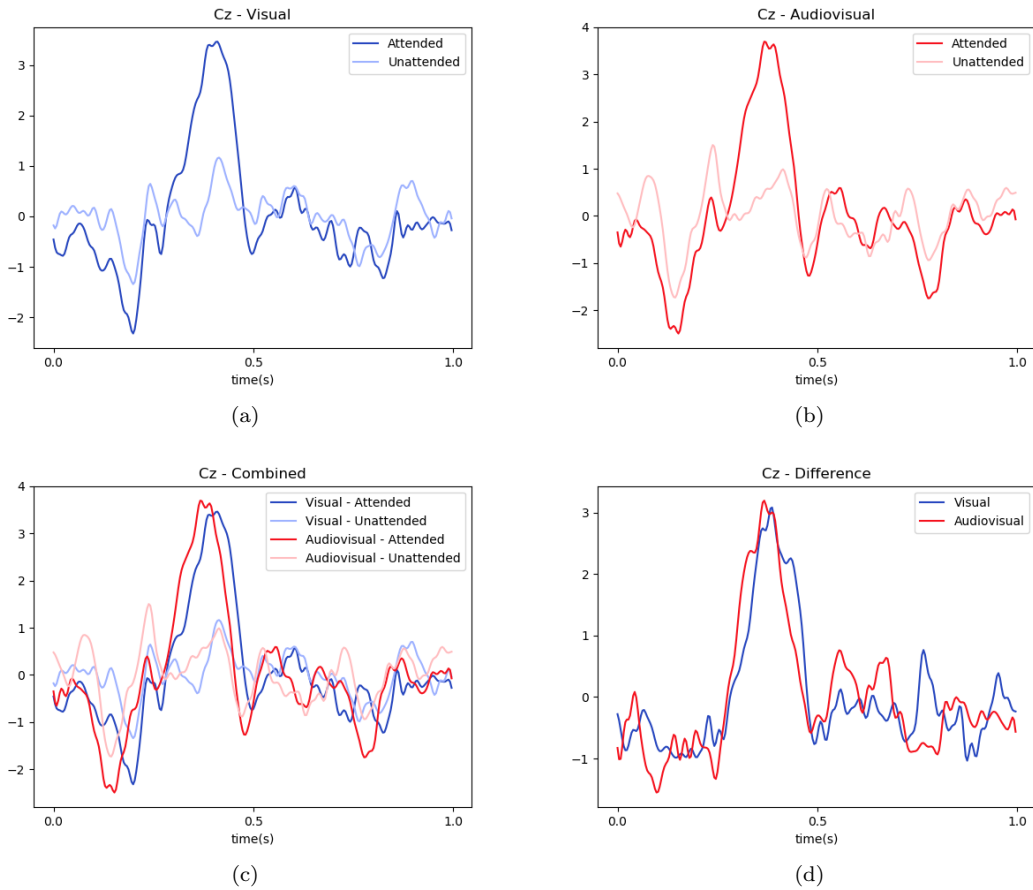


Figure 10: Recreation of Figure 9, for the grand average ERP of channel Cz. **(a-b)** Separate ERP plots (attended vs. unattended option) for each condition. **(c)** The ERP plots combined in one figure. **(d)** The difference between the attended and unattended ERP for each condition.

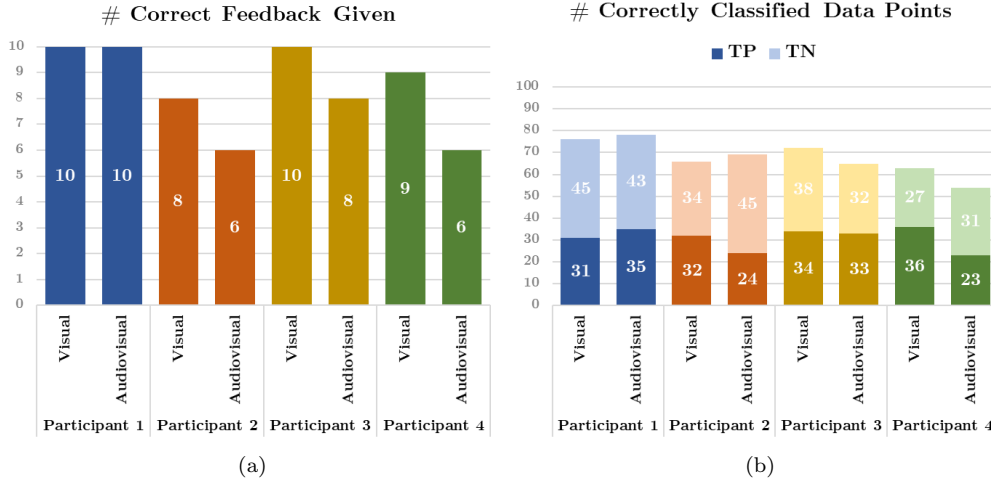


Figure 11: Results obtained from the reconstructed classifier for each participant-stimulus condition pair. (a) Number of times the predictions made by the classifier resulted in the correct feedback being given to a participant. (b) Number of data points that were correctly classified on their own. TP = True Positive, TN = True Negative.

### 3.3 Classifier Performance

Figure 11 shows the results obtained after testing the reconstructed classifier (see Section 2.6.2). Figure 11a shows the number of times correct feedback was given (with a maximum of 10, since there were 10 questions in each testing phase). Figure 11b shows the number of separate data points (brain signals resulting from “target” intensifications) that were correctly classified on their own (prediction  $\geq 0.5$  for attended option, or  $< 0.5$  for unattended option). The maximum total number of correctly classified data points was 100, as there were 10 target intensifications (and hence 10 data points) for each of the 10 questions in each training phase. The exact numbers of intensifications of the attended and unattended options were slightly different for each participant and each condition because of the randomization of the intensifications (see Section 2.3); therefore, the maximum number of true positives and true negatives was not constant across participants and conditions. However, the average maximum number of true positives, as well as true negatives, was 50.

The results of the 5-fold cross-validation (as described in Section 2.6) are shown in Figure 12. Since during this cross-validation the classifier was tested on 4 questions at a time, the maximum values discussed above should be divided by 5.

In this 5-fold cross-validation, the average mean squared error of the classifier across participants was found to be 0.194 (sd = 0.014) in the visual condition and 0.208 (sd = 0.055) in the audiovisual condition, although it is worth mentioning that the lowest average mean squared error that we encountered was in the audiovisual condition, for participant 1 (0.154). As for the number of times correct feedback was given by the classifier, we found average values of 3.750 (sd = 0.191) out of 4 questions in the visual condition and 3.200 (sd = 0.489) in the audiovisual condition. In all cases, the classifier performed worse in the audiovisual case compared to the visual case with regards to feedback (Figure 12a), even though the number of data points that were correctly classified on their own was higher in the audiovisual case for participants 1 and 2 (Figure 12b).

Finally, as described in Section 2.6.2, a classifier was trained on the data obtained from participant 1 using only the channels where we would expect a P300 response, using 5-fold cross-validation in order to obtain more reliable results. In the visual condition, this classifier gave correct feedback an average of 2.8 times (out of 4), and correctly classified an average of 23.6 data points correctly (out of 40). In the audiovisual condition, these values were 2.4 and 22, respectively.

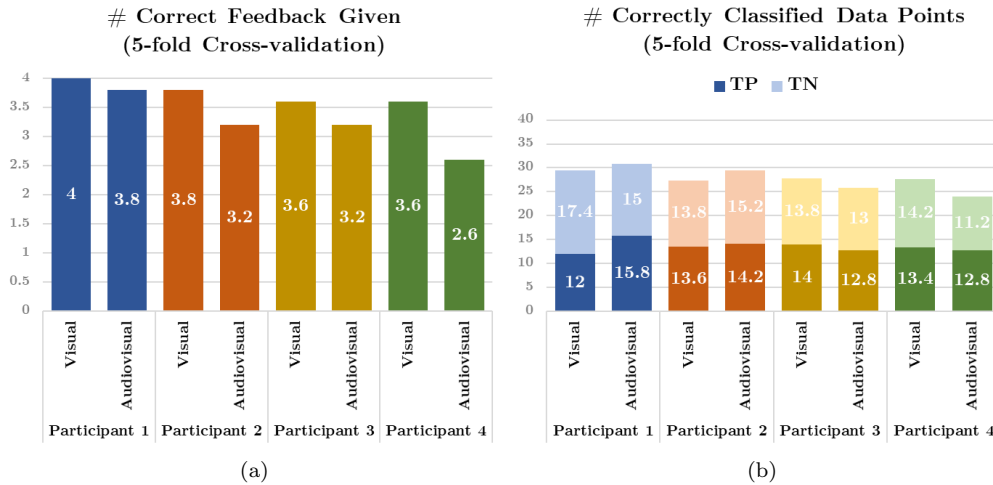


Figure 12: Results obtained from 5-fold cross-validation for each participant-stimulus condition pair (mean values). (a) Average number of times the predictions made by the classifier resulted in the correct feedback being given to a participant. (b) Average number of data points that were correctly classified on their own. TP = True Positive, TN = True Negative.

## 4 Discussion

### 4.1 Stimulus Response

In Figure 8, we can see that in both conditions, for the ERP of the attended option, the P300 component is prominently visible in the channels around the centroparietal midline. Hence, the attended target intensifications in both conditions did indeed elicit a P300 response. However, upon close inspection of Figures 9c and 9d, as well as 10c and 10d, we see that there appears to be very little difference in the stimulus response between conditions: within the margin of error of our pilot experiment, these results appear identical.

This result is also reflected in the plots shown in Appendix C: we found little to no effect of adding auditory stimuli on stimulus response, for all of the participants we tested. Hence, we did not manage to reproduce the findings by Belitski et al. (2011) in our covert attention-based interface. Since our study was only a pilot study, we did not have enough data to determine the true significance of the ERP results, so instead we turned to classifier performance: if the performance of the classifier is approximately equal in both cases, it is very unlikely that the effect of adding auditory stimuli on stimulus response is significant.

### 4.2 Classifier Performance

After examining the results of the 5-fold cross-validation we performed to test our classifier (Section 3.3; Figure 12a), we found that adding auditory stimuli to our BCI did not improve classification performance; if anything, there seemed to be a slight decrease in performance with regards to the feedback given by the classifier, although this decrease was very small. This finding was reinforced by the results obtained by looking at the predictions of single data points (Figure 12b): the differences are very small, and while for 2 out of 4 participants more data points were correctly classified in the audiovisual condition compared to the visual condition, this is reversed for the other two.

Interestingly, while on average the classifier performed worse in the audiovisual condition than in the visual condition, the lowest mean squared error we found occurred in the audiovisual condition: the data obtained from the audiovisual condition for participant 1 resulted in a mean squared error of 0.154. This result matched the opinions on the interface that were given at the end of the experiment (as mentioned in Section 2.4): participant 1 was the only participant that described the auditory stimuli as actually helpful in directing covert attention to the target intensifications. The other participants described the auditory stimuli as “distracting”.

As can be concluded from the final paragraph of Section 3.3, the classifier that was only trained on the channels around the centroparietal midline (i.e. where we expected a P300 effect) performed worse than the classifier trained on all channels, but still performed above chance. Hence, our classifier probably does use the P300 component, but it is unlikely that it *only* uses this component (otherwise, the performance of the classifier using only the channels around the centroparietal midline should be closer to that of the classifier that uses all channels). This result is in line with the results of the experiment by Treder and Blankertz (2010), who found that, besides P300, several other ERP components such as the P2 component were also modulated in the ERP-based Hex-o-Spell interface.

### 4.3 General Discussion

As the experiment described in this paper was only a pilot study, we did not gather enough data in order to conclusively state whether adding auditory stimuli to our visual covert attention-based BCI had an effect on stimulus response or classifier performance. The limited amount of data we obtained does not appear promising, but future research would be needed to see if there truly is no effect.

Since we used only two options, the classifier already performed well above chance in the visual condition: for participant 1, it even gave

correct feedback for all 10 questions in the testing phase (see Figure 11a). The effect of auditory stimuli might be more pronounced if a BCI with more than two options is used (such as the ERP-based Hex-o-Spell interface), since performance in the visual case would likely be worse; future research is necessary to determine if this is indeed the case.

Even though we chose our auditory stimuli so that they were pleasant to listen to for prolonged periods of time, several participants noted that they found the beeps to be distracting. We do not know exactly why this is the case; it may be due to there being beeps during both target and distractor intensifications (which we described and justified in section 2.3) or due to the use of earphones as opposed to speakers. More research would be needed to determine how these stimuli can be made less distracting.

Despite our findings, we believe that the topic of improving the performance of covert attention-based BCIs by using auditory stimuli warrants further investigation, in particular to make covert attention-based BCI spellers (such as the ERP-based Hex-o-Spell interface) more clinically viable. The ability to control BCI spellers without eye movements would allow a larger portion of their target population, such as people with locked-in syndrome or ALS, to use them with less effort, which would enable them to communicate with the outside world in a more pleasant way.

## References

- Belitski, A., Farquhar, J., & Desain, P. (2011). P300 audio-visual speller. *Journal of Neural Engineering*, 8(2), 025022.
- Blankertz, B., Dornhege, G., Krauledat, M., Schröder, M., Williamson, J., Murray-Smith, R., & Müller, K. R. (2006). The Berlin Brain-Computer Interface presents the novel mental typewriter Hex-o-Spell.
- Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., & Schalk, G. (2010). Does the ‘P300’ speller depend on eye gaze?. *Journal of neural engineering*, 7(5), 056013.
- Buffer BCI. (2019). Retrieved from [https://github.com/jadref/buffer\\_bci](https://github.com/jadref/buffer_bci)
- Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6), 510-523.
- Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S. & Desain, P. (2009). The brain-computer interface cycle. *Journal of neural engineering*, 6(4), 041001.
- Guger, C., Harkam, W., Hertnaes, C., & Pfurtscheller, G. (1999, November). Prosthetic control by an EEG-based brain-computer interface (BCI). In *Proc. aaate 5th european conference for the advancement of assistive technology* (pp. 3-6).

Krusiński, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. *Journal of neuroscience methods*, 167(1), 15-21.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118(10), 2128-2148.

Scikit-learn. (2019). `sklearn.linear_model.RidgeCV`. Retrieved from [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.RidgeCV.html#sklearn.linear\\_model.RidgeCV](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html#sklearn.linear_model.RidgeCV)

Treder, M. S., & Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and brain functions*, 6(1), 28.

Wood, C. C., Allison, T., Goff, W. R., Williamson, P. D., & Spencer, D. D. (1980). On the neural origin of P300 in man. *In Progress in brain research* (Vol. 54, pp. 51-56). Elsevier.

## Appendix

### A. Question list

Below, we list all of the questions that were used in our experiment. These questions were chosen to be easily decidable for the participants in our experiment (who were all Dutch). The questions on the left are those to which the correct answer is “yes”; the correct answer to those on the right is “no”; in both cases, the questions are listed in alphabetical order.

- |  |   |
|--|---|
| 1. Are needles sharp?                            | 1. Are lamps used to tell time?               |
| 2. Can a lamp emit light?                        | 2. Are ovens used to freeze things?           |
| 3. Can a microwave heat things up?               | 3. Are roses tall trees?                      |
| 4. Can a pen be used to write things?            | 4. Are there 50 minutes in an hour?           |
| 5. Can you sit in a chair?                       | 5. Are trees usually purple?                  |
| 6. Can you watch a movie on a television?        | 6. Can you watch a movie on a radio?          |
| 7. Do books contain words?                       | 7. Do cats live in the ocean?                 |
| 8. Do dolphins live in the ocean?                | 8. Does 1 plus 1 equal 3?                     |
| 9. Does 1 plus 2 equal 3?                        | 9. Does 4 minus 4 equal 3?                    |
| 10. Does the sun usually rise every day?         | 10. Does a circle have corners?               |
| 11. Does wood come from trees?                   | 11. Does a square have five corners?          |
| 12. Is a mouse an animal?                        | 12. Do humans have wings?                     |
| 13. Is a violin a musical instrument?            | 13. Do pigs walk on two feet?                 |
| 14. Is a window transparent?                     | 14. Is a calculator a musical instrument?     |
| 15. Is bread edible?                             | 15. Is a fork used to cut things?             |
| 16. Is drinking water good for you?              | 16. Is Amsterdam the capital of Germany?      |
| 17. Is football a sport?                         | 17. Is a table an animal?                     |
| 18. Is London the capital of the United Kingdom? | 18. Is Berlin the capital of the Netherlands? |
| 19. Is sugar sweet?                              | 19. Is drinking poison good for you?          |
| 20. Is water wet?                                | 20. Is fire cold?                             |

## B. Experiment Explanation

The text below was given to the participants beforehand, to inform them about the experiment. This does not include the consent form, which was part of a separate document.

Brain-Computer Interfaces (BCIs) can be used to communicate with the outside world without using your peripheral nervous system, which for example can be used by people with certain conditions that prevent them from communicating in conventional ways. However, these people sometimes also have trouble moving their eyes, which is why covert attention-based BCIs exist: they can be used without users having to move their eyes to direct overt attention to stimuli. However, these covert attention-based BCIs perform worse than regular BCIs. In this Bachelor Thesis an attempt is made to improve their performance by combining the visual BCI with auditory stimuli (beeps).

## Experiment

The experiment you're going to take part in consists of two parts, which will be explained below; you will be told the order in which you encounter these parts before the experiment begins. Both parts start with a short text. Read this text carefully, and once you are done, press any button on the button box in front of you to continue.

### Visual Part

The visual part of the experiment consists of a training and a testing phase. During the training phase, you are tasked with answering 10 simple yes/no questions, with a break of 5 seconds after every 2<sup>nd</sup> question. Answering a single question goes as follows:

- A question will appear on screen; you have a few seconds to read it, and decide whether the answer is yes or no.

- Afterwards, two options (YES and NO) will appear left and right of the center of the screen, respectively, as well as a red cross in the center (the fixation cross). **Please keep your eyes on the red fixation cross whenever it is shown: moving your eyes (or blinking) may result in bad trails. Moving your eyes and blinking is allowed whenever the fixation cross is not shown.**
- After a second, the options will start to “intensify”, meaning that they increase in size and rapidly return to their original size, in random order. For most of the intensifications, the color of the options will remain identical (white), but sometimes an option will turn red instead. Your task is to **count the number of times your chosen answer turns red, without moving your eyes away from the fixation cross; ignore the “regular” (white) intensifications!**

During the training phase, no feedback will be given as to what answer the BCI thinks you chose, as it is still gathering data to calibrate a classifier based on your brain. After training is done, a message will pop up asking you to wait for the experimenter to finish training the classifier; as soon as this is done, you will be met with another text for you to read (press a button on the button box to proceed). After this, the testing phase starts: the procedure here is the exact same as for the training phase (10 questions, with a break after every 2<sup>nd</sup> question), only after the intensifications have stopped you will receive feedback on what the classifier thinks your answer was (the selected answer turns blue, while the other turns gray).

## Audiovisual Part

The audiovisual part is almost identical to the visual part, only with the addition of auditory beeps (which you will hear via earphones). During a “regular” intensification (that is, when an option increases in size but remains white), a 400 Hz beep is played in the ear on the same side as the intensification (i.e. left for YES, right for NO). During the “red” intensifications (which you have to count again for your chosen option), a 500 Hz (that is, slightly higher) beep is played in the corresponding ear (also for the option you did not choose). You can use these beeps as additional information to count the number of times your chosen option turns red.

Besides this, the audiovisual part is identical to the visual part; there is a training and a testing phase, both of which consist of 10 questions, with a break after every 2<sup>nd</sup> question, and with feedback only being provided during the testing phase.

## Personal Data

During this experiment, EEG data is recorded from your brain. This data is stored for use in this Bachelor Thesis, but your name or any other personal data will not be linked to any of this data, only your participant number will be used (see the start of this explanation). Before the experiment, you will be given a consent form in which you will have to fill in some personal data, but this will not be linked to the experimental data.

**If you have any remaining questions, feel free to ask them!**

## C. ERP Plots

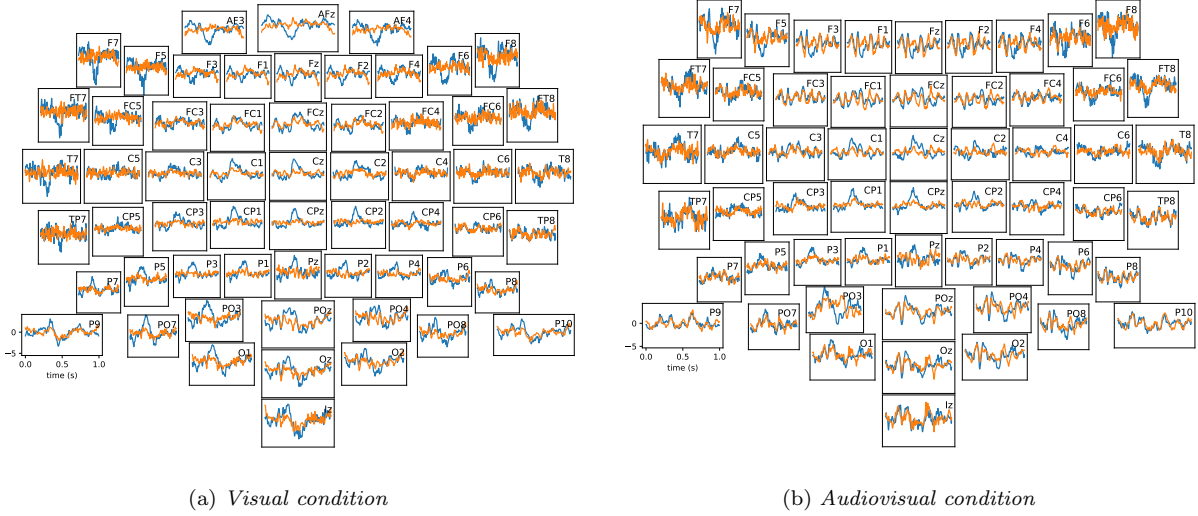


Figure 13: Recreation of Figure 8, for the data obtained from participant 2. Blue lines correspond to intensifications of the attended option; orange lines to those of the unattended option. Note that bad channels were left out of these plots.

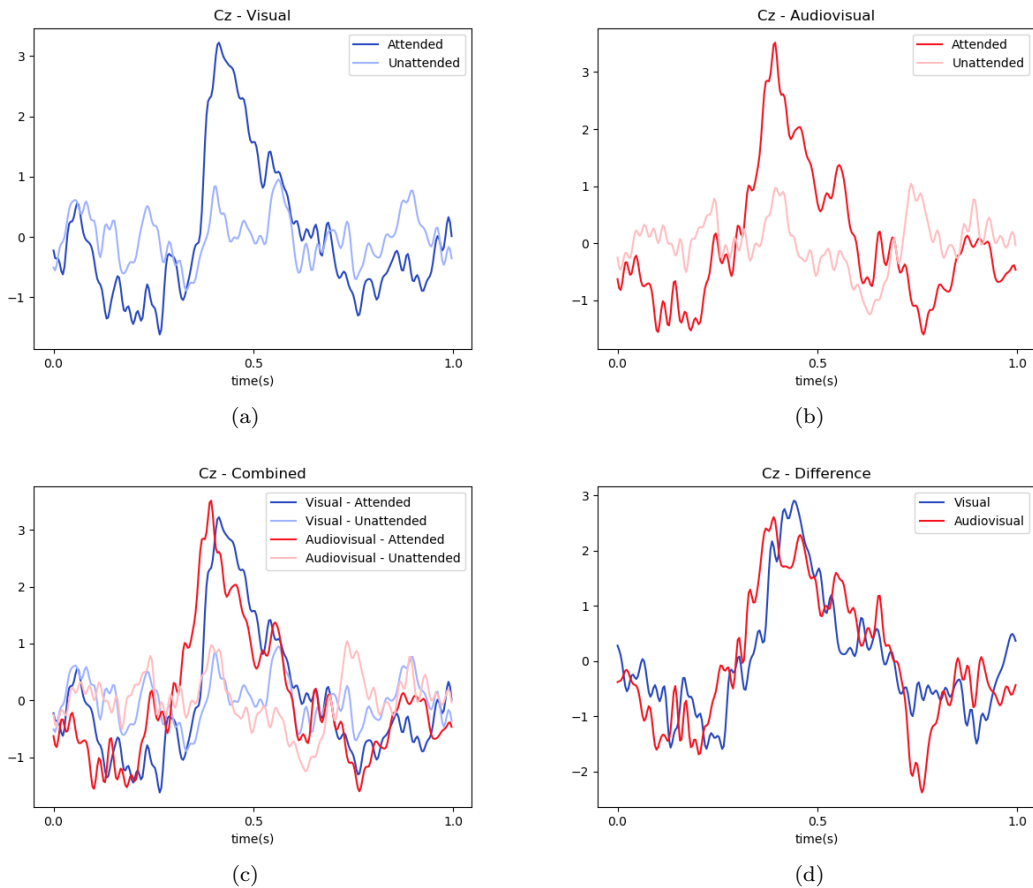
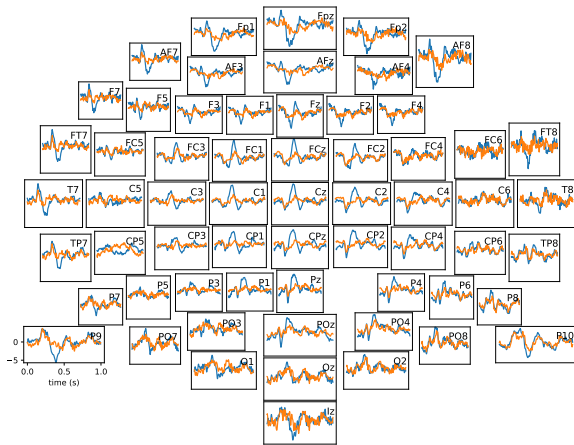
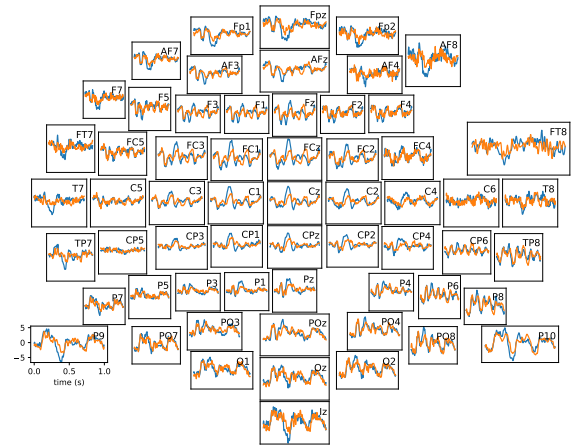


Figure 14: Recreation of Figure 9, for the data obtained from participant 2. (a-b) Separate ERP plots (attended vs. unattended option) for each condition. (c) The ERP plots combined in one figure. (d) The difference between the attended and unattended ERP for each condition.

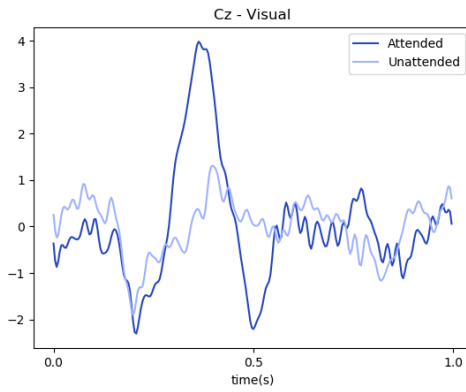


(a) Visual condition

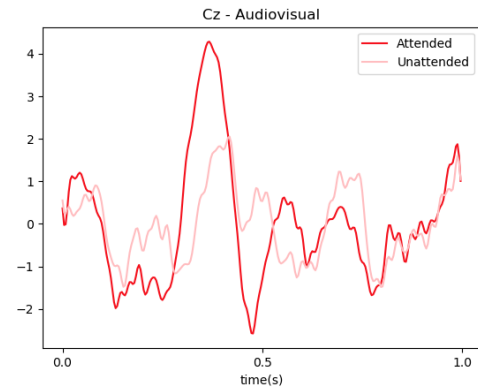


(b) Audiovisual condition

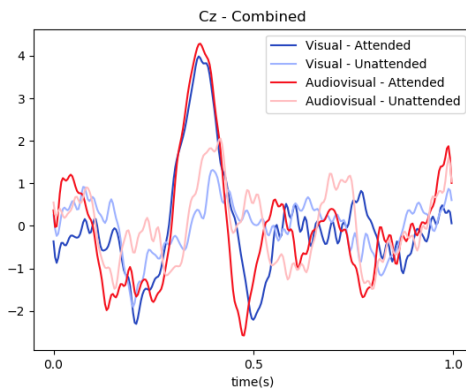
Figure 15: Recreation of Figure 8, for the data obtained from participant 3. Blue lines correspond to intensifications of the attended option; orange lines to those of the unattended option. Note that bad channels were left out of these plots.



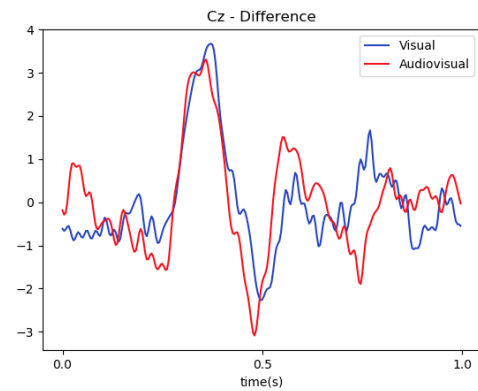
(a)



(b)

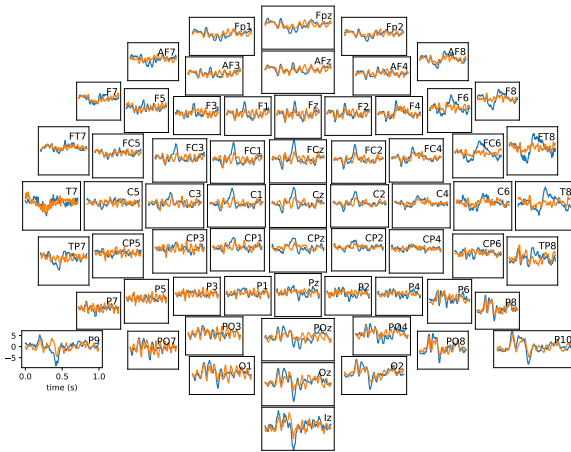


(c)

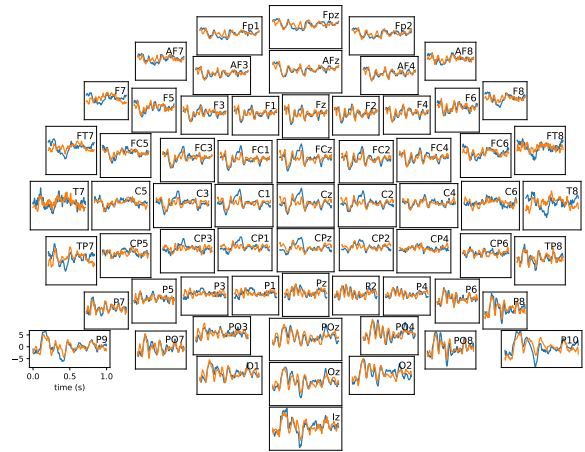


(d)

Figure 16: Recreation of Figure 9, for the data obtained from participant 3. (a-b) Separate ERP plots (attended vs. unattended option) for each condition. (c) The ERP plots combined in one figure. (d) The difference between the attended and unattended ERP for each condition.

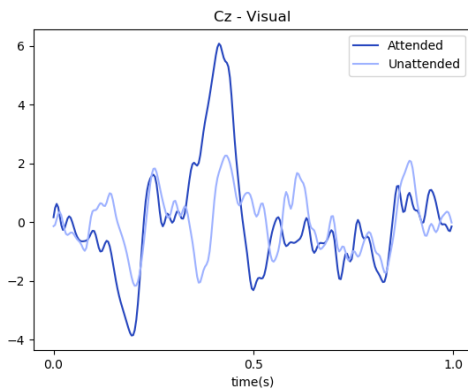


(a) Visual condition

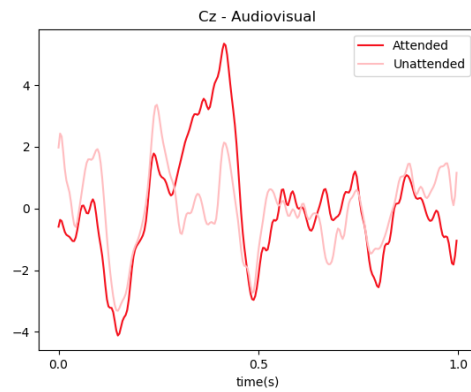


(b) Audiovisual condition

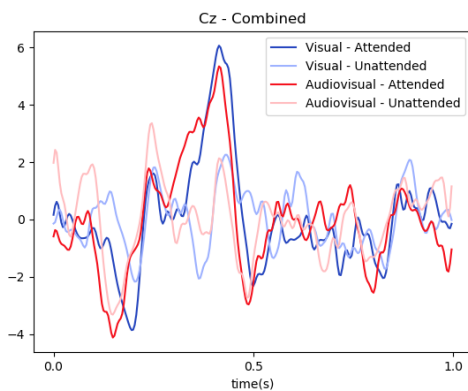
Figure 17: Recreation of Figure 8, for the data obtained from participant 4. Blue lines correspond to intensifications of the attended option; orange lines to those of the unattended option. Note that bad channels were left out of these plots.



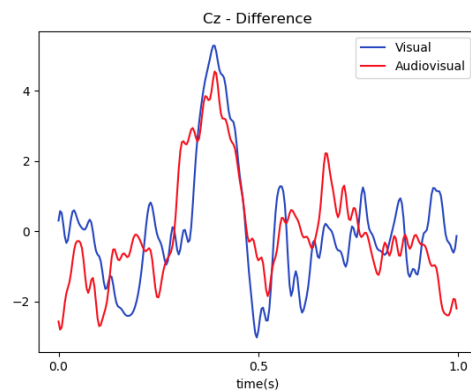
(a)



(b)

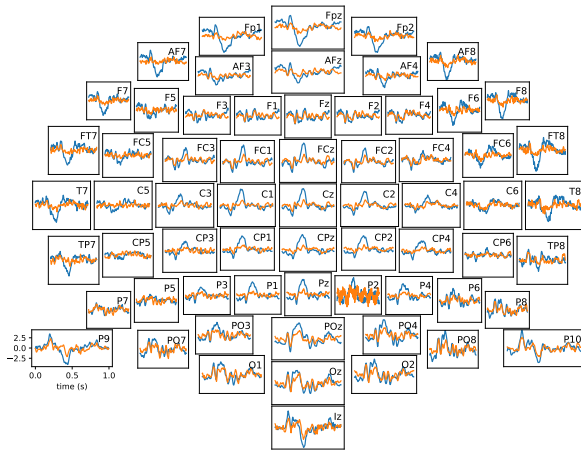


(c)

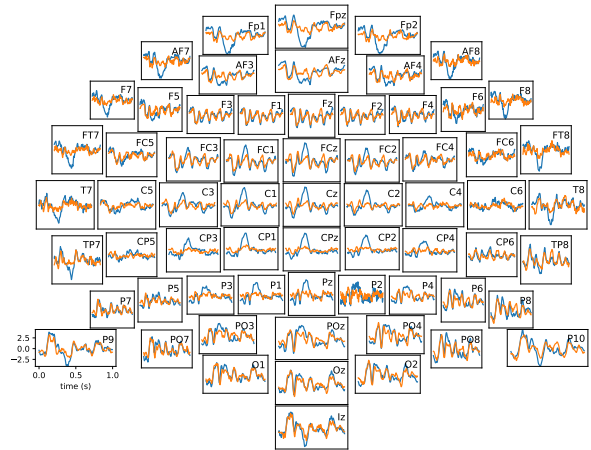


(d)

Figure 18: Recreation of Figure 9, for the data obtained from participant 4. (a-b) Separate ERP plots (attended vs. unattended option) for each condition. (c) The ERP plots combined in one figure. (d) The difference between the attended and unattended ERP for each condition.



(a) Visual condition



(b) Audiovisual condition

Figure 19: Recreation of Figure 8, for the grand average ERP. Blue lines correspond to intensifications of the attended option; orange lines to those of the unattended option. In these plots, bad channels were not left out, as these are averages of multiple participants.