Using Error Potentials to Improve Auditory BCI

Author:
J.E.M. Muris (Jill)
Student Number: s4787579

Supervisor:
C.S. Verbaarschot (Ceci)[1]

Reading Committee:
C.S. Verbaarschot (Ceci)[1] &
dr. F.A. Grootjen (Franc)[2]

[1] Donders Centre for Cognition; Donders Institute for Brain, Cognition and Behavior; Radboud University Nijmegen
[2] Radboud University Nijmegen

Bachelor's Thesis in Artificial Intelligence

Radboud Universiteit Nijmegen

Artificial Intelligence
Faculty of Social Sciences
Radboud University
Nijmegen, the Netherlands
July 9th, 2019

Using Error Potentials to Improve Auditory BCI

J.E.M. Muris and C.S. Verbaarschot

July 9th, 2019

Abstract

Ever since the start of the development of brain-computer interfaces (BCIs), they have been envisioned as a tool that could help impaired people in their interaction with the world. BCIs translate brain signals into computer commands. Auditory BCIs, which only require an intact hearing, could be used as a communication tool for patients with locked-in syndrome who have very little control over their muscles or none at all, making it hard to control BCIs based on (eye) movements. In this paper we execute a pilot study in which we propose a way to improve the accuracy of an auditory BCI. In order to answer a binary question, subjects control the BCI by changing their state of mind as a result of a shift of attention between auditory streams. A certain component of the measured brain signal, an error potential, could be detected during the feedback phase of the BCI in case of incorrect provided feedback. If an error potential could be detected, the initial feedback could be corrected, increasing the overall accuracy of the BCI. The results from this experiment show that there was no noticeable improvement in the performance of the auditory BCI after addition of a second brain signal (error potentials); the overall accuracy even seemed to diminish.

1 Introduction

A brain-computer interface (BCI) forges a direct online connection between brain and machine. Van Gerven and his colleagues (2009) introduced the concept of a BCI cycle in which they explain the stages of BCI research. This type of research starts with a user who has to perform a certain task. Brain signals are measured during the execution of this task. Different types of measurement methods can be used for this. One frequently used non-invasive method is electroencephalography (EEG) which measures the evoked electrical current potentials of neuronal changes using electrodes that are placed on the scalp. BCI can also be based on e.g. induced changes in the power of certain frequencies, which are captured in ERSPs (event-related spectral potentials). In the next stage of the cycle, the collected data is preprocessed, relevant features are extracted and the system makes a prediction about the intention of the user. This prediction is given as feedback in a predetermined modality (e.g. visual, auditory or motor). BCIs can be used online or offline. In an online BCI predictions are derived from the measured data and fed back to the user in real-time, whereas in an offline BCI, predictions are made after the experiment has ended and the user does not receive immediate feedback.

One of the most well-known examples of a BCI are visual spellers. One famous application of this is the matrix speller as introduced by Farwell and Donchin (1988). In such a speller, letters or possible answers are arranged in the shape of a matrix. During the use of such a speller each of the rows is intensified in a random order, followed by intensifications of each of the columns in a random order. The selected element is in the intersection of a row and column and has a unique intensification pattern. This unique pattern can be retrieved from the measured data and a prediction of the selected element can be made. The brain signal that is used in such spellers is called a P300 (Polich, 2007). This is a component of an ERP (event-related potential), which is a time-locked signal generated in response to a stimulus. The P300 potential occurs when a deviant stimulus is presented in a sequence of normal stimuli. In the case of the matrix speller an oddball is presented when the selected element flashes. The ERP is measured at around 250-500 ms after the onset of the deviant stimuli and appears as a positive deflection (Figure 1). The downside of this type of spellers is that for an effective implementation, the user has to be able to actively control and move their eye muscles. Some users that could benefit from the use of BCI spellers, might not be able to do so for this reason.

A possible solution to this problem are auditory BCI. For an auditory BCI a user only has to have normal hearing, i.e. the user should not have an hearing impairment. Hillyard and his colleagues (1973) stood at the frontier of research in this field. They presented their participants with two different auditory streams, one to each ear. Each stream consisted of 2 tones: a lower "non-target" and higher "target" frequency tone. The signals were played at random, creating an oddball paradigm. The users were asked to focus on one of the two streams and silently count the targets in that stream. The counted tones elicited a P300 potential. The sequence of P300 potentials could be matched to the sequence of tones occurring in the stimuli to retrieve the attended auditory stream. This outcome proves that the P300 oddball paradigm can be applied to the auditory modality, which makes this concept applicable in the field of BCI. The downside of these types of auditory P300-based BCIs is their poor accuracy, in comparison to their visual counterpart. The accuracy that can be achieved using a P300-based visual speller can be as high as 94.62% whereas the accuracy of P300-based spellers using the auditory modality lies at around 65% (Furdea et al., 2009). Hill, Lal, Bierig, Birbaumer and Schölkopf (2005) predicted an accuracy of 82.4% in an offline study on auditory BCI; Hill and Schölkopf (2012) found an accuracy of 84.8% in an online version of this study.

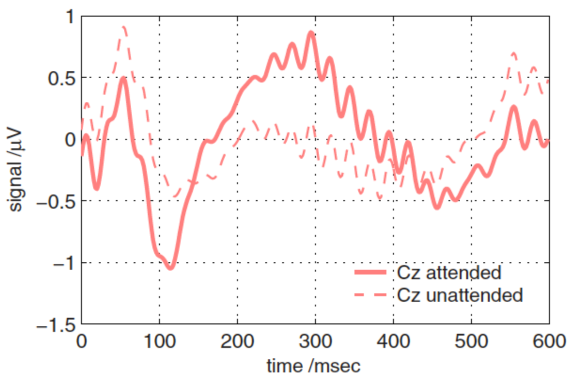


Figure 1: *EEG signal at electrode Cz. The line of the attended condition represents a P300 potential. From Hill & Schölkopf (2012)*

BCIs can be used as a system to make changes in the world by translating brain signals into commands that could be executed by an external device. As a result of this, BCIs can help people communicate that are not able to do so efficiently in the conventional ways. These people could control a BCI by adjusting their brain signals as a result of e.g. specific (eye) movements (Birbaumer, 2005). This does not offer a solution to people that have a more severe case of paralysis, e.g. in an advanced stage of ALS (amytrophic lateral sclerosis). The stage these people are in is called completely locked-in state (CLIS) if they have no rudimentary control of any muscle, including the eye muscles, and locked-in state (LIS) if they have control over at least one muscle (Smith & Delargy, 2005). Auditory BCIs require normal hearing and normal EEG and therefore it could be used as an alternative method that enables people in such stage to communicate with the outside world. An improvement in performance of such BCI is beneficial for this type of users and therefore should be aimed for.

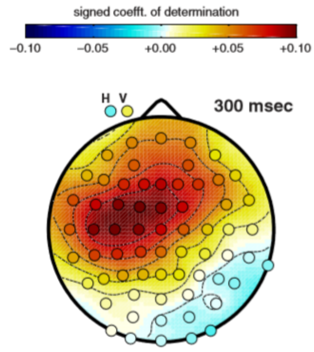


Figure 2: *Scalp map showing the spatial distribution of a P300 potential. From Hill & Schölkopf (2012)*

Figure 2 shows the spatial distribution of a P300 potential across the scalp of a participant in the experiment by Hill and Schölkopf (2012).

A second component of the ERP called error potential (ErrP) could be introduced to improve the overall performance of P300-based BCIs. Earlier research found that ErrPs can be used to improve accuracy of a P300-based visual BCI (Spüler et al., 2012). Zeyl, Yin, Keightley and Chau (2016) proposed that ErrP could also be used in the auditory modality to increase performance. An error potential presents itself after erroneous feedback, and therefore could be used to correct this initial incorrect feedback, leading to a higher overall BCI accuracy. Ferrez and del R. Millán (2008) coined the term "interaction ErrP" to refer to cases in which error potentials were generated during an human-computer interaction. These interaction ErrPs consist of four consecutive peaks (shown in Figure 3): a positive peak at 200 ms, a negative peak at 250 ms, a positive peak at 320 ms, and a last negative peak at 450 ms. Especially the two last peaks are salient. The peaks are focussed at the central parts of the scalp, with the first peak being more fronto-central (Figure 4).

In this paper, we will perform a study that is intended as a proof of concept, where we try to investigate whether addition of ErrPs improves the performance of a P300-based auditory BCI. The specifics of the auditory stimulus used in the BCI are adapted from the paper by Hill and Schölkopf (2012). Participants will be presented with two different auditory streams. Each stream corresponds to an answer to a trivial binary question, i.e. left corresponds to 'yes' and right corresponds to 'no'. Each participant is asked to silently count the number of target (oddball) beeps on the side that is associated with the answer he or she intends to give, eliciting a distinct P300 sequence. A classifier is used to recognize the P300 potentials and derive on which stream the user focussed. The prediction from this classifier is fed back to the user in real-time. During the feedback phase, a second linear classifier is active to see whether an error potential occurred. If this is the case, then the feedback from the first classifier is corrected. The accuracy of the second 'ErrP' classifier should be above chance to be able to improve the overall performance. We expect that addition of a second classifier that is active during the feedback phase will improve the performance of the BCI.

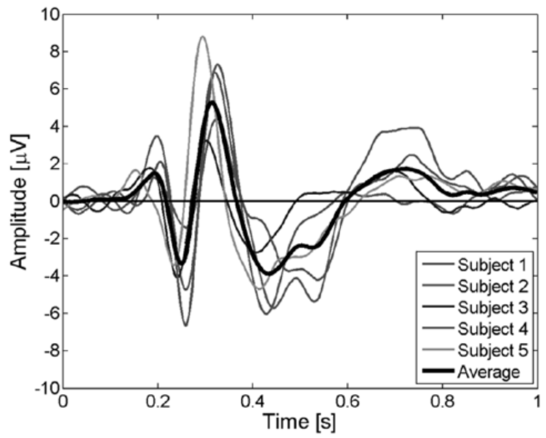


Figure 3: *Average EEG for the difference error-minus-correct at channel FCz plus grand average. From Ferrez & del R. Millán (2008).*

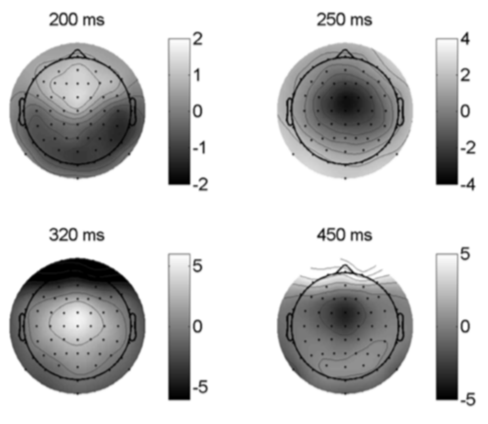


Figure 4: *Scalp potential topographies at the time stamps of the peaks from Figure 3. From Ferrez & del R. Millán (2008).*

2 Methods

2.1 Subjects

Three healthy participants (all male) with an age range of 20-21 years took part in the pilot experiment. All subjects Dutch, but were proficient in the English language. None of the subjects had a history of significant hearing defects or were diagnosed with dyslexia. They all gave informed consent before taking part in this experiment.

2.2 Stimuli and Task Design

Each trial started with an English binary question and ended with an answer to this question. We created a total of 60 binary questions (40 of which were created by a fellow thesis group-student), of which we assumed they were common knowledge. All questions are reported in Appendix A. The order in which the questions appeared to the participants was random and the block of 60 questions was repeated if all questions had occurred once; as a result, each question was posed in 4 trials. To confirm that the subjects knew the correct answers to the questions, we let them fill in a questionnaire, containing all 60 presented questions, after they took part in the experiment.

The subjects received both visual and auditory stimuli. All 240 trials had the same lay-out visually from the perspective of the participant. In Section 2.2.1 we will discuss the visual stimulus and in section 2.2.2 we will discuss the auditory stimulus in more detail.

2.2.1 Visual Stimuli

First of all, the participant is shown a welcome message, in which he/she is asked to press a button whenever he/she is ready to start the experiment. The experiment starts by showing a binary question in white font on a dark grey background. The question is shown for 3 seconds, which is sufficient for the participant to carefully read the question and come up with an answer. After this, a yellow fixation cross appears in the center of the screen along with the two possible answers to the question: 'yes' and 'no'. Each answer is assigned to a fixed side of the screen, i.e. 'yes' is always shown on the left, and 'no' is always shown on the right. Each answer is fixed to a side to remind the participant which auditory stream is associated with which answer, i.e. the left stream is associated with 'yes' and the right stream is associated with 'no'. In the provided instructions the participant is asked to keep his/her gaze focussed on the fixation cross to minimize the influence of eye artifact. While this screen is shown, the auditory streams are played. The display is shown in a total of 4.5 seconds. In the next screen feedback is provided to the subject: one of the two possible answer disappears, leaving the answer to the question on the screen. The feedback is shown for 2 seconds. During the feedback, the fixation cross also remains on the screen so the subject should try to not move his/her eyes away from this. The subject should still be able to see the feedback in the corner of his/her eye. An overview of the sequence of visual presentations is presented in Figure 5. The visual part of one trial ends with a break message and a break of 5 seconds. During the break no fixation cross is shown, so the participant does not have to remain focussed. The next trial starts with a new question. The participant has a longer self-paced break after each 10 questions and an experimenter-paced break after each 60 questions.

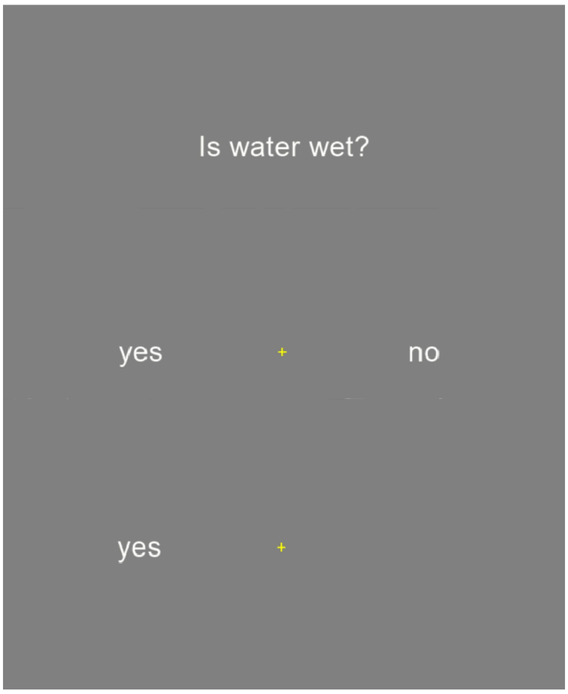


Figure 5: *Visual stimulus representation. The order in which the screens are shown to the participant is from top to bottom.*

2.2.2 Auditory Stimuli

The auditory stimulus specifics are identical to the ones used in a paper by Hill et al. (2005). The auditory stimulus consists of two distinct auditory streams. The setting at which the streams are presented is as follows: the participant is sat on a chair wearing an EEG cap on and he is looking at the fixation cross on the screen. One stream is played from a speaker that is placed to the left of the screen, the other stream is played from a speaker to the right of the screen, both directed towards the participant. The streams consist of 50 ms long square-wave beeps. The specifics of the two auditory streams are displayed in Table 1. Both streams consist of 2 different beeps of different frequencies. The lower frequency beep is the non-target beep and the higher frequency beep is the target beep. The non-target beep is played more frequently and therefore the target beep will be an oddball and elicit a P300 potential. Each auditory stream starts with 3 non-target beeps, such that the participant can familiarize him-/herself with the sound and has time to lock his/her attention to the correct stream. After this, each new tone that is being played has a probability of 0.3 of being a target beep. In case no target beep has been played yet but the auditory stream only has one tone left to play, we forced this tone to be a target beep, since each trial should have at least one target beep in each stream. The left auditory stream starts playing 70 ms after the right auditory stream started to make sure that both streams, with different periods, finish at the same time, and to make it easier for the participant to distinguish both streams.

Each stream is associated with an answer to the visually proposed question. The participant is asked to solely focus on the stream that is associated with his desired answer. To enlarge the possibility of a P300 potential, the user is asked to silently count the target beeps in his/her attended stream.

	LEFT	RIGHT
Frequency non-target (Hz)	800	1500
Frequency target (Hz)	880	1650
Number of beeps	7	8
Period (ms)	555	490

Table 1: *Specifics of the two auditory streams.*

2.3 Data Acquisition

Stimulus presentation was created and presented using PsychoPy version 3 (Peirce et al., 2019). Anaconda Spyder version 3.1.4 (Anaconda, 2017) was used for all programming related to the training and testing of the Scikit Learn linear RidgeCV-classifiers (Scikit-learn, 2019). For EEG collection we used a Biosemi headset (Biosemi, 2019) with 64 active AgCl electrodes along with 4 electrodes for EOG measurement (1 placed above and 1 under the left eye, and 2 on each side of the head) and 2 mastoid electrodes, sampled at 256 Hz. Retrieved data is saved to and retrieved from a Buffer BCI toolbox (Farquhar, 2014) created for a master's BCI course at the Radboud University.

2.4 Data Analysis

We analyzed the data both online and offline. The experiment was divided into a training and a testing phase. Both phases consisted of 120 trials (questions), with a total of 240 trials. There was no difference in stimuli between the two phases. The participant is made to believe that in all trials the answer to the question is derived from the collected EEG data, however this is not the case in the training phase. For the actual predictions made by the system during the test phase, we used two linear classifiers (Scikit-learn, 2019). We expect to see P300 potentials during the presentation of the target beeps in the auditory stimulus and error potentials during the presentation of incorrect feedback. Therefore we trained one linear classifier, referred to as the 'P300 classifier', on data collected during the auditory stimulus and the second linear classifier, referred to as the 'ErrP classifier', on data collected during the feedback phase.

In the training phase we made use of a 'fake classifier'. The 'classifier' provided feedback to the participant to keep the presentation of feedback consistent between the training and testing phases. We programmed it to deliberately make incorrect predictions in 25% of the trials such that the classifier always received a consistent number of data containing an actual ErrP. This error percentage was based on earlier research in which a selection accuracy of 75% was used on average for artificially generated feedback in similar research (Zeyl, Yin, Keightley & Chau, 2016).

2.4.1 Online

During the execution of the experiment we collected 600 ms long slices of recorded EEG data after presentation of a tone from either stream, both target and non-target. This resulted in 9 slices of 600 ms per trial (4 for left auditory stream, 5 from the right stream), with a total of 2160 slices over the course of the complete experiment. We do not save data from the first 3 tones from both streams since they are always non-targets and they would cause a bias in our analysis. This data is used later to train and test the linear 'P300' classifier on. We also collected 1 second slices of recorded EEG data for the ErrP classifier. These slices were made after the feedback was presented to the participant.

After the training phase (i.e. the first 120 trials) ended, we trained the two classifiers on the collected data. The training of the ErrP classifier is straightforward: we perform some standard pre-processing steps. Firstly, we detrend the data followed by bad-channel removal, application of a common average reference spatial filter, and a spectral bandpass filter of 1-10 Hz. We end with bad-trial removal before finally training the linear classifier.

The training of the P300 classifier is more complex. The first 3 preprocessing steps are the same as those used for the ErrP classifier. In the fourth step we use a bandpass filter of 0.1-45 Hz. We compute the averages over all tones from the left stream and all tones from the right stream such that we end up with 2 averages in the form of a channel-by-sample matrix. We process these averages by subtracting the left mean values from the right mean values. We compute and save these differences for all 120 trials, and use them as training data for the P300 classifier. We omit bad-trial removal in the preprocessing of the data for this classifier because we do not want to have biases in our computed averages.

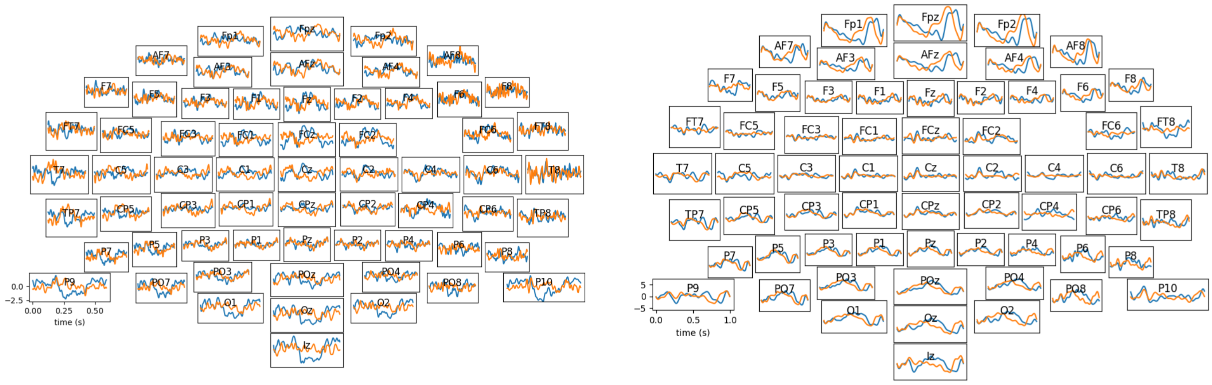
In the trials of the testing phase, we sliced the same pieces of recorded EEG data as we did for the training phase. We used the same preprocessing steps as during the training phase on the data as we performed on the data from the training phase. The P300 classifier returns 9 predictions, namely one after each received slice. We compute two averages: one over the 4 slices that were made after presentation of tones from the left auditory stream, and one over the 5 slices from the right auditory stream. The side with the highest average is chosen as the final prediction. This prediction is then fed back to the participant. The ErrP classifier makes one prediction, namely on the data sliced after the presentation of the P300 classifier. From this single prediction we could make a correction to the feedback of the P300 classifier, in case of detection of an ErrP. If an ErrP was detected, the participant received a message stating that the feedback that had been shown earlier appeared to be incorrect.

2.4.2 Offline

We saved all slices of data that were collected during the execution of the experiment and used these to perform an offline analysis. This analysis was similar to the online analysis procedure, except that it was done using data from the 64 electrodes from the cap without the EOG or mastoids electrodes. The same preprocessing steps were used.

We tested the newly trained classifiers on the slices of data retrieved during the test phase of the online experiment to see whether the predictions made by the offline classifiers were the same as those obtained from the online classifiers. We used the new predictions to compute the accuracy of the offline classifiers, which will be reported in the Results section of this paper, alongside the accuracy of the online classifiers.

We used the preprocessed data to make plots for every one of the 64 electrodes used for the EEG



(a) (b)

Figure 6: *The scalp potential topographies created from data generated from the EEG of subject 1. (a) P300 potential. At time 0, a tone from an auditory stream is played. The blue and orange lines represent the attended and unattended auditory streams, respectively. (b) Error potential. At time 0, feedback is presented to the participant. The blue and orange lines represent the trials in which erroneous and correct feedback was provided, respectively.*

recording. In Section 3.1 we zoom in on a selection of these plots. We make a comparison between the plot created from data from the Cz electrode and FCz electrode and Figure 1 from Hill and Schölkopf (2012) and Figure 3 from Ferrez and del R. Millán (2008). We also created scalp potential topographies for every participant for both the P300 and the ErrP data.

We computed the overall accuracy of the P300 and ErrP classifiers. We made a distinction between the accuracy obtained during the online experiment and the accuracy obtained during the offline analysis. The overall accuracy percentages were computed by dividing the number of trials with correct feedback by the total number of trial in the testing phase. Feedback was considered correct if it was the intended answer to the question. The accuracy percentages for the P300 classifiers were computed as follows: we retrieved the number of trials in which the classifier made a correct prediction and divided this by the total amount of trials in the testing phase of the experiment (i.e. 120 trials). To be able to compute the accuracy of the ErrP classifiers we first retrieved the number of trials in which the classifier detected an error potential. This is the denominator of the accuracy fraction later. We also needed to know the number of trials in which the ErrP classifier correctly detected an error potential (the true positive trials). The accuracy was then computed dividing the number of true positive trials by the number of trials an ErrP was detected. To end up with percentages all outcomes were multiplied by 100.

3 Results

In this section we summarize the results. Interpretations and conclusions can be found in Section 4.

We started the analysis of the results with evaluating the questionnaire. The questionnaire contained all 60 yes/no questions. All three participants filled in the correct answers to the questions and therefore we can assume in our analysis that the subjects intended to give the correct answer to the question in each trial. Hence, in our analysis, we can use the correct answer to the questions as being the answers of the subjects.

3.1 Event-Related Potentials Plots

We recreated the ERP signals that were recorded during the EEG recordings. We created scalp potential topographies of all subjects for both the P300 data and the ErrP data. To be able to make comparisons with Figure 1 and 3 we zoom in on electrode Cz and FCz, respectively. In this section we only report the plots from the first participant. Similar plots generated using data from the other participants are reported in Appendix B.

3.1.1 P300

In Figure 6a we plotted the scalp potential topography generated from the P300 data. We see a positive peak at around 300 ms in the central region of the scalp. Next, we made a plot of electrode Cz (Figure 7). The plot has the the following shape: first the line for the unattended condition is above the line for the attended condition, after which the lines cross and the attended line has a peak after which they cross again finishing with the attended condition line below the unattended condition line. The blue attended line shows a positive peak at around 300 ms.

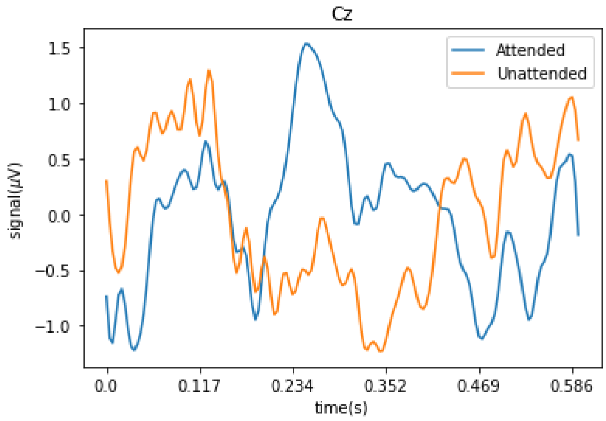


Figure 7: *EEG plot of electrode Cz generated using the slices of data made during the presentation of the auditory stimulus from participant 1.*

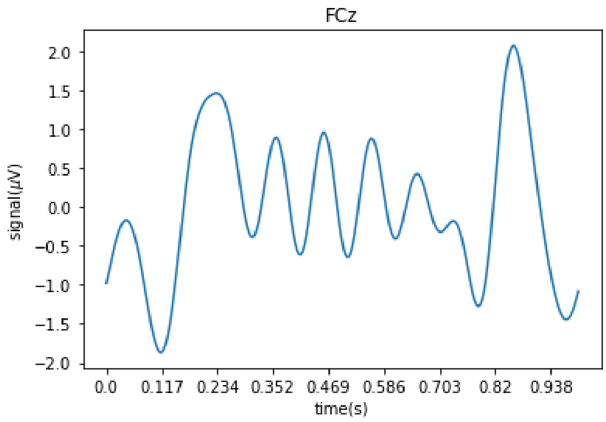


Figure 8: *The error-minus-correct plot generated of data from subject 1.*

3.1.2 Error Potential

In Figure 6b we reported the ErrP scalp potential topography we generated from the training and test data combined from subject 1. The frontal plots shows a salient positive peak at around 700 ms, while the occiptal plots show a negative peak at the same timestamp. Next, we zoom in on the plot from electrode FCz (Figure 8). As is common in the reporting of error potentials, we opted for a presentation of the error-minus-correct differences.

We also created a plot showing the error-minus-correct lines from all subjects, along with a grand average (Figure 9). The line from participant 2 seems to have peaks with higher amplitudes.

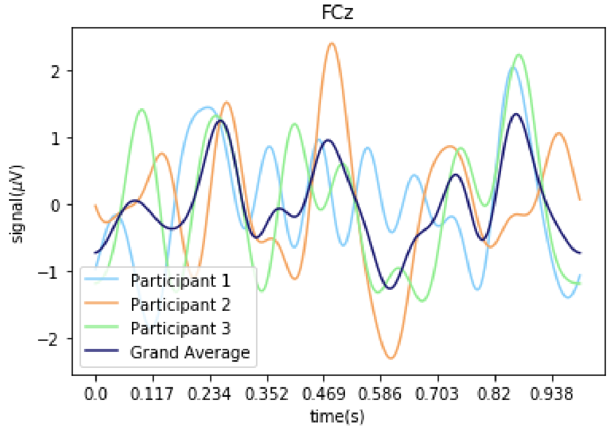


Figure 9: *Error-minus-correct from electrode FCz for all participants along with a grand average.*

3.2 Classifier Accuracy

The classifiers make the classification based on the recorded EEG data and provide the feedback to the user. We computed the accuracy of both the classifiers that were trained online and offline. In the first section we report the overall BCI accuracy; in later sections we discuss the P300 and ErrP classifier performances in more detail.

3.2.1 Overall

In the computation of the overall performance, we made a distinction between the online and offline classifiers. At the end we want to know whether addition of the ErrP classifier improves overall accuracy. To be able to draw a conclusion, we report two accuracy percentages per set of classifiers, computed on data from the test phase: the accuracy obtained by using only the P300 classifier and the accuracy obtained by using the P300 classifier in combination with corrections from the ErrP classifier. The results are reported in Table 2.

3.2.2 P300

The calculated accuracy percentages of the P300 classifiers are presented in Table 3. These percentages were also presented in Table 2, but for convenience we presented them along with the number of correct trials in this new table. The average accuracy of both classifiers lies at around 50%, which means that both the classifiers perform at chance, i.e. the classifiers do not make estimated predictions as to which stream was attended.

Overall performance		**Subject 1**	**Subject 2**	**Subject 3**
Online	Only P300 classifier	58.33%	51.6%	45.83%
	Incl. corrections from ErrP classifier	54.17%	50%	55%
Offline	Only P300 classifier	45.83%	60%	40%
	Incl. corrections from ErrP classifier	46.67%	34.17%	3.33%

Table 2: *The overall performance of the online and offline classifiers. We make a distinction between the performance of the classifiers from the online and offline analysis.*

3.2.3 ErrP

In the final section of the Results we discuss the performance of the two ErrP classifiers. We present the achieved accuracy percentages in Table 4. These data resulted in the reported accuracy percentages. In almost all cases the classifier detected more ErrPs than there should have been when assuming that an ErrP only occurs in trials with incorrect initial feedback. The accuracy of the online classifier was higher than the accuracy of the offline classifier. The overall accuracy of both classifiers combined was below chance.

		Correct	Accuracy
Online	S1	70	58.33%
	S2	62	51.6%
	S3	55	45.83%
			51.93%
Offline	S1	55	45.83%
	S2	72	60%
	S3	48	40%
			48.61%
			50.27%

Table 3: *Accuracy percentages for both the online and offline P300 classifiers. Subjects are referred to as S1, S2 and S3. The 'Correct' column represents the trials in which the P300 gave correct feedback.*

		# ErrP detected	TP	Accuracy
Online	S1	21	8	38.10%
	S2	28	13	46.43%
	S3	17	14	82.35%
			Average	55.63%
Offline	S1	43	22	51.16%
	S2	51	10	19.61%
	S3	52	4	7.69%
			Average	26.15%
			Overall Accuracy	40.89%

Table 4: *Correctness percentages computed for both the online and offline ErrP classifiers. The subjects are referred to as S1, S2 and S3. TP = true positive.*

4 Discussion

We performed a study in which we tested whether error potentials could be used to improve auditory BCI. Since we did a pilot experiment, it is not possible to draw general conclusions. We do try to explain our results and offer possible suggestions for future research in the next paragraphs.

We make comparisons between our P300 plots of subject 1 and the plots from Hill and Schölkopf (2012). We see that the topographical plot of our subject (Figure 6a) has the same spatial distribution at 300 ms as Figure 2. The zoomed-in plot of the Cz electrode (Figure 7) shows some resemblance to Figure 1, we can conclude that our auditory BCI did elicit a P300 potential in subject 1. However, when looking at the plots from the other two participants, which are reported in Appendix B, we can see that the P300 potential was not as prominent for these participants. Based on these 3 participants alone, we cannot draw general conclusions with respect to the occurrence of a P300 potential as a result of this specific auditory set-up. Further research, possibly with more subjects, would be required to draw such conclusions.

Since we expect to find an ErrP that is similar to the interaction ErrP from Ferrez & del R. Millán (2008), we compare our results with Figures 3 and 4. When comparing the topographical plot from subject 1 (Figure 6b) with Figure 4, we see little resemblance. We make a more precise comparison by comparing the zoomed-in error-minus-correct plot of the FCz of subject 1 (Figure 8) and the grand average from Figure 3. We are not able to spot the characteristics of an 'intention' ErrP (Ferrez & del R. Millán, 2008) in our Figure 8. The error-minus-correct grand average line generated using data from the all subjects (Figure 9) also did not show resemblance with the components of the interaction ErrP. Hence, we can conclude that the feedback phase of this experiment did not elicit an interaction error potential. A possible explanation for this observation is that the participants noticed that in the first trials of the experiment no real classifier was used, which caused a weaker response to the feedback and bad training data for the ErrP classifier.

For future research, it should be stressed to subjects that during all trials the feedback is generated using their brain signals. A second explanation for the lack of error potentials could be the simple design of the interface. The feedback might need to be provided in a more noticeable manner, by for example intensifying the correct answer rather than simply removing the answer that the P300 classifier predicted was not the one chosen by the user.

The weak P300 potentials could explain the low accuracy of the online and offline trained 'P300' classifiers. Both classifiers performed at chance and therefore made random predictions. The 'ErrP' classifiers performed well below chance. This could be explained by the fact that there were no error potentials elicited during the feedback phase which can be seen in the event-related potentials plots (e.g. Figure 9). The classifiers did correct some mistakes from the P300 classifier, but also frequently corrected trials for which the provided feedback was already correct, reducing overall accuracy. In some cases (offline subject 1 and online subject 3) a higher accuracy was found in the condition with corrections of the ErrP classifier. The only reason we can think of as to why this occurred is nothing more than chance. The low performance of the classifiers could also be explained by the type of classifier that was used. We chose simple linear classifiers, whereas this data might need more complex classifiers that could adapt more easily to the peculiarities of the recorded EEG data. This could be tested in future research.

To conclude our research, we were not able to confirm nor deny our research question because of the small scale. The concept of a second classifier trained to spot interaction error potentials and as a result corrects earlier classification mistakes still sounds promising. For future research, more thought should be put into the design of the interface and the choice of classifier.

References

Anaconda Spyder package 3.1.4 (2017). Spyder: The Scientific Python Development Environment. Retrieved from `https://anaconda.org/anaconda/spyder`

Biosemi (2019). Retrieved from `https://www.biosemi.com/`

Birbaumer, N. (2006). Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6), 517-532.

Farquhar, J. (2014). Buffer BCI toolbox. Retrieved from `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.

Ferrez, P. W., & Millán, J. D. R. (2008). Error-related EEG potentials generated during simulated brain–computer interaction. *IEEE transactions on biomedical engineering*, 55(3), 923-929.

Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., & Kübler, A. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3), 617-625.

Hill, N. J., Lal, T. N., Bierig, K., Birbaumer, N., & Schölkopf, B. (2005). An auditory paradigm for brain-computer interfaces. *In Advances in neural information processing systems* (pp. 569-576).

Hill, N. J., & Schölkopf, B. (2012). An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli. *Journal of neural engineering*, 9(2), 026011.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108), 177-180.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E. & Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

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`

Smith, E., & Delargy, M. (2005). Locked-in syndrome. *Bmj*, 330(7488), 406-409.

Spüler, M., Bensch, M., Kleih, S., Rosenstiel, W., Bogdan, M., & Kübler, A. (2012). Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. *Clinical Neurophysiology*, 123(7), 1328-1337.

Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., ... & Desain, P. (2009). The brain–computer interface cycle. *Journal of neural engineering*, 6(4), 041001.

Zeyl, T., Yin, E., Keightley, M., & Chau, T. (2016). Improving bit rate in an auditory BCI: Exploiting error-related potentials. *Brain-Computer Interfaces*, 3(2), 75-87.

Appendix A

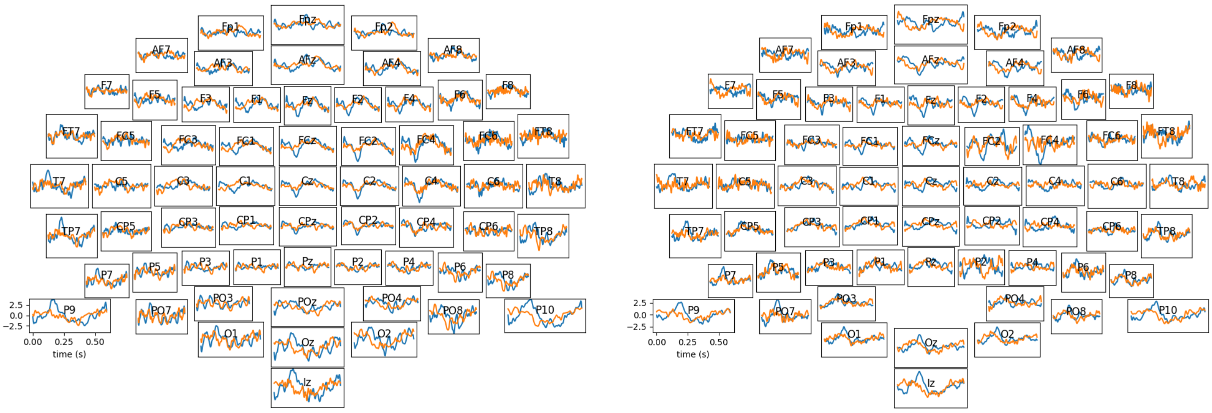
The trivial Yes/No Questions Used in the Experiment

1. Is it January 5th today?
2. Are the Netherlands and China neighbour countries?
3. Is Nijmegen the capital of the Netherlands?
4. Does everyone have the same DNA?
5. Does the Netherlands have a king?
6. Is the earth flat?
7. Do you use a knife to eat soup with?
8. Do you need fire to light a candle?
9. Does an airplane fly faster than a bird?
10. Do bees sting?
11. Is Paris the capital of France?
12. Does a circle have corners?
13. Does 1 plus 2 equal 3?
14. Are roses tall trees?
15. Are lamps used to tell time?
16. Is a violin a musical instrument?
17. Is Berlin the capital of the Netherlands?
18. Are ovens used to freeze things?
19. Are pillows usually soft?
20. Do pigs walk on two feet?
21. Is a window transparent?
22. Is a table an animal?
23. Is water wet?
24. Can a microwave heat things up?
25. Is Amsterdam the capital of the Netherlands?
26. Do plants need water to survive?
27. Can you watch a movie on a television?
28. Are there 50 minutes in an hour?
29. Do cats live in the ocean?
30. Is bread edible?
31. Are needles sharp?
32. Is grass usually green?
33. Does the sun usually rise every day?
34. Is London the capital of the United Kingdom?
35. Is drinking poison good for you?
36. Does 4 minus 4 equal 3?
37. Are clouds white?
38. Do dolphins live in the ocean?
39. Can a lamp emit light?
40. Does a square have five corners?
41. Is a fork used to cut things?
42. Are tulips flowers?
43. Is a calculator a musical instrument?
44. Is sugar sweet?
45. Are trees usually purple?
46. Can you sit in a chair?
47. Is drinking bleach good for you?
48. Does 2 times 4 equal 8?
49. Should you ride a bicycle on a highway?
50. Can a pen be used to write things?
51. Is a mouse an animal?
52. Does wood come from trees?
53. Can you watch a movie on a radio?
54. Is football a sport?
55. Is an eraser used to write?
56. Is nail polish used on nails?
57. Is your birthday twice a year?
58. Are phones used to make calls with?
59. Does the Netherlands have a lot of mountains?
60. Do cars only drive backwards?

Appendix B

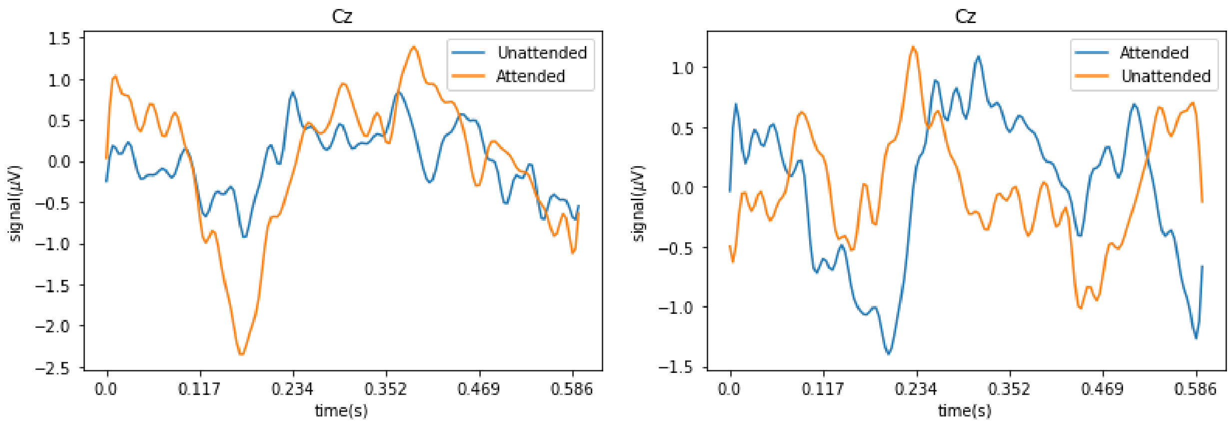
Event-Related Potential Plots of Subjects 2 and 3

B.1 P300



(a) (b)

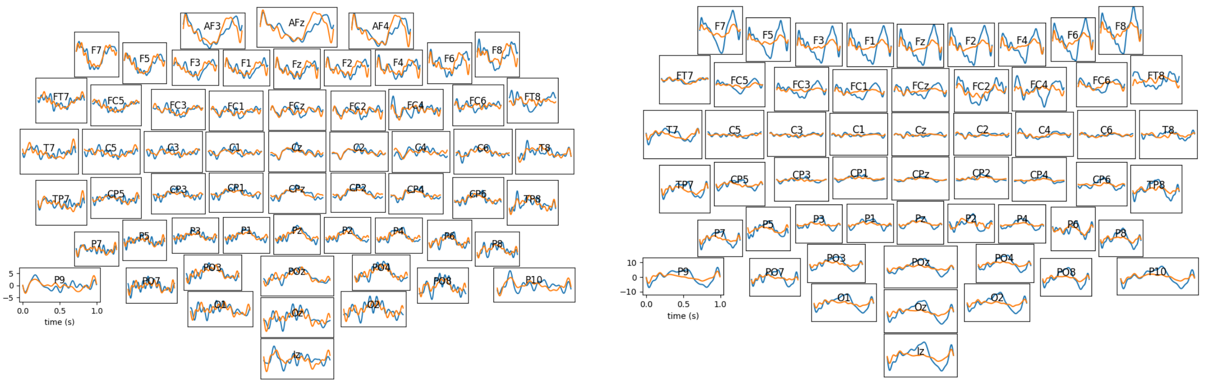
Figure 10: *Scalp potential topography created using P300 data from (a) subject 2 and (b) subject 3. Blue and orange lines resemble the attended and unattended streams, respectively.*



(a) (b)

Figure 11: *Zoomed plot of the Cz electrode generated using P300 data from (a) subject 2 and (b) subject 3.*

B.2 ErrP



(a) (b)

Figure 12: *Scalp potential topography created using ErrP data from (a) subject 2 and (b) subject 3. Blue and orange lines resemble the trial in which erroneous and correct feedback was given, respectively.*

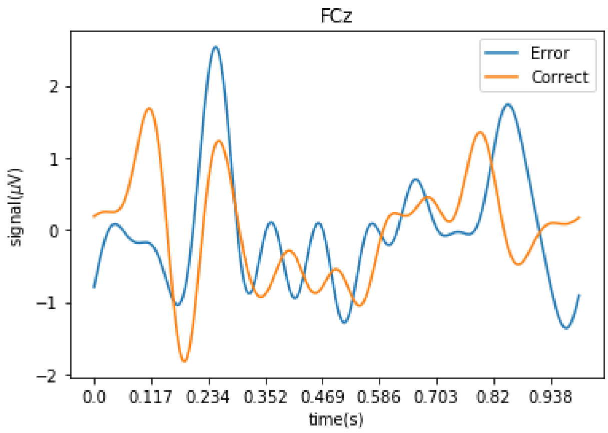
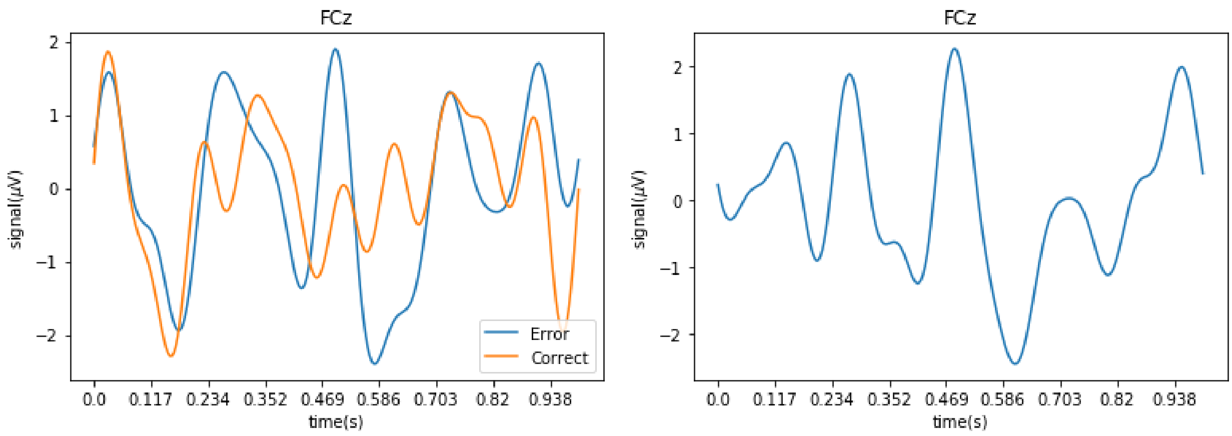
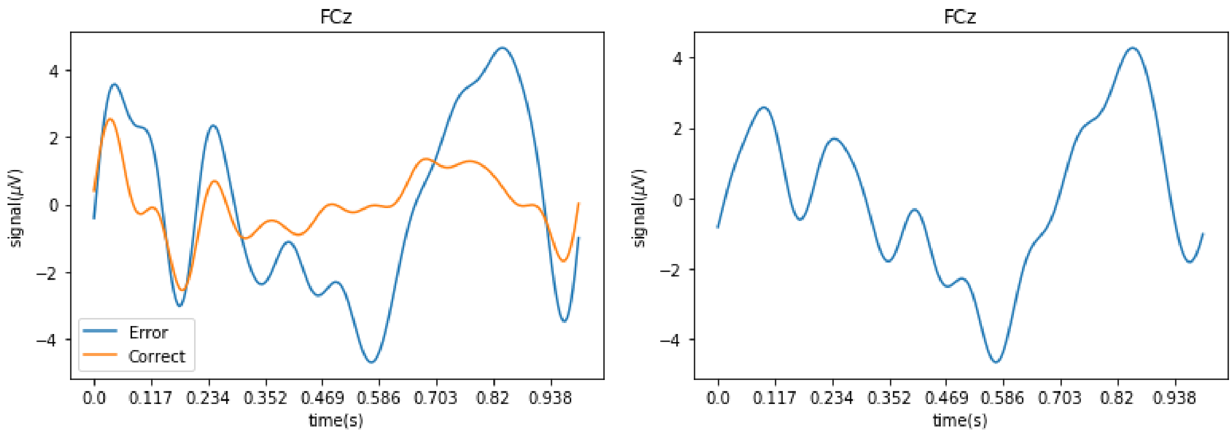


Figure 13: *The 'Error' line represents the trials in which an ErrP was detected, whereas the 'Correct' lines detect the lines where no ErrP was detected. Generated of data from subject 1.*



(a) (b)

Figure 14: *Zoomed in plot of electrode FCz generated using data from subject 2. (a) This plot makes a distinction between the trials in which erroneous and correct feedback was given. (b) This plot shows the error-minus-correct line.*



(a) (b)

Figure 15: *Zoomed in plot of electrode FCz generated using data from subject 3. (a) This plot makes a distinction between the trials in which erroneous and correct feedback was given. (b) This plot shows the error-minus-correct line.*