Measuring vigilant attention: Predictive power of EEG derived measures on reaction time, subjective state and task performance

Master thesis

Frank Klein Douwel (S9413332)

University Of Twente

Date: February, 2016

Supervisors: (1) Dr. Rob van der Lubbe, (2) Dr. Martin Schmettow

Department: Cognitive Psychology and Ergonomics

Abstract

EEG is considered a promising and sensitive measure of sustained attention. Commonly used EEG analysis methods are Event Related Potentials (ERP), Fast Fourier Transformation (FFT) and Event-Related De-synchronization (ERD). The current study examined which of these methods yielded the most useful EEG derived measure for predicting reaction time, subjective state and task performance. It was expected that ERD would yield the most sensitive measure because it takes both the time and frequency domain into account.

Participants performed a monotonous visual discrimination task and were instructed to respond only to infrequent target stimuli. During the task EEG was recorded, the level of drowsiness was scored subjectively and the behavioral responses to the target- and non-target stimuli were registered. Different EEG measures were derived from the same signals by performing FFT, ERP and ERD analysis. Separate regression analyses were performed for each EEG method as predictor of subjective scores, reaction time and task performance. The EEG derived measures were statistically compared by using information criteria. It was found that FFT Alpha band located at parietal sites proved to be the most informative predictor for task performance and reaction time. For subjective state, ERP, located at the occipital site was the most informative predictor. Based on the results from this study, it could not be confirmed that ERD yielded the most predictive EEG measure.

Introduction

Modern working environments are highly automated and this has changed the role of operators to system supervisors (Berka et al., 2007). The work often involves passive monitoring of screens over prolonged periods of time and the operator only needs to react when critical events occur. Because these events are rare and unpredictable it is mentally demanding to stay vigilant (Hancock, 2012), but assessing this state is not so clear-cut (Smit, 2004). Currently, EEG is considered to be a promising measure. It has not only a high temporal accuracy and resolution but recent findings suggests it also has predictive power. Because different EEG methods are used in vigilance research, their usefulness needs to be examined in order to exploit the full potential of this measure.

During monotonous tasks it is observed that almost inevitably vigilance decline sets in (Warm et al., 2008). For example, in WW II it was observed that on submarine radar control, misses of enemy vehicles occurred especially at the end of a watch. A failure to detect crucial or critical signals can lead to severe accidents and therefore research focuses on human factors influencing vigilance (Nickerson, 1992) and its causes and mechanisms (Warm et al., 2008).

Despite of sixty years of research, vigilance and its mechanisms are still poorly understood and we still don't know how to identify vigilant workers (Finomore, Matthews, Shaw, & Warm, 2009) or influence the vigilance state. A possible explanation for this is that vigilance has proven to be a broad and multi-dimensional concept with no consensus on how it is defined and measured (Langner & Eickhoff, 2013). A general definition is: "The human capacity to endogenously maintain focused attention on a portion of the environment or a certain task to monitor for changes over prolonged periods of time" (Martel, Dähne, & Blankertz, 2014). It is still debated upon if 'maintaining focused attention' refers to the availability of internal resources, the process of self-regulation or staying engaged to the task and also the capacity to stay focused can vary largely between individuals. The definition does not specify which task or environment is meant, so this can refer to a broad range of applied settings with varying conditions. Determining vigilance state is therefore not univocal (Smit, 2004) and depends strongly on the definition and scope of the research.

As mentioned in the beginning, most monitoring and screening tasks consist of long periods of inactivity. Because this is perceived as unchallenging and non-rewarding, it can easily lead to under-arousal and task disengagement, causing lapses of attention (Langner & Eickhof, 2013). Because critical signals are rare, sustained attention is needed. Paradoxally, this is perceived as mentally demanding (Warm, Dember, & Hancock, 1996) because it requires effort and can be stressful (Warm et al., 2008). Therefore this is even more likely to set in a decrement in performance. Research on sustained attention is ongoing, but the phenomenon is poorly understood and it is difficult to find a sensitive measure for it (Robertson & Garavan, 2004).

To examine how attention is influenced during monotonous tasks, several vigilance tasks have been designed like the Psychomotor Vigilance Task (PVT; Lim & Dinges, 2008), the Continous Performance Task (CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956) and the Mackworth Clock Test (Mackworth, 1948). Each task and its measures have their shortcomings and limitations. The PVT involves the detection of an unwarned stimulus to which participants have to respond as fast as possible. A limitation is that the response to the same signal is used, so no distinction is made between critical and non-critical stimuli. Therefore it measures only the readiness for speeded responding to unwarned stimulation (Langner & Eickhoff, 2013). The CPT and Mackworth Clock test involve discrimination between target stimuli and non-target stimuli. Both stimuli appear intermixed in a constant sequential stream consisting of regular intervals at a fixed rate. Participants need to respond to targets and withhold a response to non-targets over prolonged periods of time. In general, the stream consists of more non-targets than targets and only the targets occur infrequently and unpredictable. What it measures is a person's sustained and selective attention. Like the PVT, this can be derived from the speed of responses, but also from the number of detected and missed targets and the number of discrimination errors made. A shortcoming is that most visual discrimination tasks contain stimuli-events at a regular and high rate. This does not quite resemble real working environments in which the event-rate is at a slower pace and stimuli occur irregular. Another shortcoming is that performance on the task is retrospective which does not serve the purpose of preventing accidents from happening.

Next to the behavioral measures, also subjective measures are used. Participants are asked to assess their level of alertness, fatigue or drowsiness on a questionnaire. Subjective judgment is considered a sensitive measure of vigilance (Smit, 2004) and it involves awareness and judgment of a participant's own internal state. This is derived from the physical activation level and an estimation of availability of resources (Matthews & Davies, 2001). A disadvantage of subjective measuring is that, during monotonous tasks, participants might not be well able to rate their own internal state due to under-arousal and drowsiness. Furthermore participants have to be interrupted, which can interfere with executing the task. Like task performance, subjective state is also a retrospective measure.

The aim of the current study is to assess sustained attention during monotonous tasks by using Electro-encephalography (EEG) and to investigate which analysis method yields the most sensitive and predictive EEG derived measure. To account for the shortcomings and limitations of the aforementioned tasks and measures, a new visual discrimination task will be used which contains very few targets and less frequent and irregular occurring non-target events. Because critical target stimuli are rare in this setting, vigilant attention can only be partly derived from the behavioral responses to target events. Therefore, Electro-encephalography (EEG) will be used, especially during non-critical target events. EEG is useful because it is a direct measure of the internal state which records electrical activity of the brain on the entire scalp. The recorded brain-activity consists of time-varying differences in voltage over different parts of the brain which are thought to reflect neural activity (Cohen, 2014). Contrary to subjective and performance measures it is also continuous over prolonged periods and it does not interrupt with the task. EEG is a sensitive measure because it can detect changes in real-time (Blankertz et al., 2010) and predict errors and lapses of attention (Martel et al., 2014). Results of EEG research could therefore be used to identify vigilant workers or drive the development of new BCI applications that use EEG to monitor and alert operators in real time.

Related to sustained attention, most EEG studies conducted use Event-related Potentials (ERP's) and Fast Fourier Transformation (FFT) as analysis methods. Both methods can be determined relatively quickly and easy, but an important limitation is that each focuses on either the temporal or spectral domain. An ERP is an averaged brain-response and looks at when events happen. Because an ERP is time-locked to an event, it has no information from the spectral domain. Several ERP studies have been conducted while performing a Go/No-go task. In a Go/No-go task participants need to respond to frequently occurring non-target stimuli and withhold a response to rare target stimuli. Zordan, Sarlo and Stablum (1997) used a Go/No-Go task with a random instead of fixed stream of stimuli. The No-Go trials revealed a frontal P3 and the Go trials a smaller posterior P3b, which also had a shorter latency. A P3, or late P300 component, has a positive amplitude ranging from 250 to 500 ms and is in general elicited by rare stimuli. It was concluded that the more posterior P3 recorded in the Go trials was related to the processing of task-relevant properties of the stimulus, while the more frontal P3 recorded in the No-Go trials was an index of inhibitory processes. In concordance with earlier findings, this suggests that the P3 response is associated with the attentional system which can involve both the top-down attentional enhancement of stimuli

(Hopfinger & West, 2006) as well as resource allocation of attention (Wickens, Kramer, Vanasse, & Donchin, 1983). The P3 amplitude is sensitive to the amount of resources and visual attention deployed to the task and lower P3 peak values are indicative of lapses of attention which can lead more easily to behavioral errors (O' Connell, Dockree, Robertson, Bellgrove, Foxe, & Kelly, 2009).

Next to ERPs, Fast Fourier Transformation (FFT) is often used as an EEG analysis method. FFT characterizes an EEG signal by decomposing it in its frequency components. It basically looks at which frequencies occur and when they occur. Frequencies are thought to represent brain-dynamics and they consist of periodic oscillations that are the result of complex interactions of neuronal networks (Cohen, 2014). A relevant FFT study conducted by Martel et al. (2014) investigated if the detection of target stimuli could be predicted from FFT. A monotonous visual discrimination task was used consisting of regular and frequently occurring non-target stimuli. Participants had to respond to irregular target stimuli only. It was found that activity in the Alpha band (8–14 Hz) increased over parietal-occipital area, and gradually accumulated 10 seconds before a missed target. Alpha activity preceding successfully detected target stimuli was lower and this decrease was thought to be associated with decreased concentration. Next to changes in the Alpha band, the Theta band has also proven to be a sensitive measure of sustained attention. Smit (2004) focused on differences in Theta band changes between a simple discrimination task and a demanding task involving working memory. It was found that during both tasks Theta activity increased above baseline. However, no difference in mental effort was found between both tasks. Because FFT mainly looks at the spectral domain, it does not have an accurate temporal resolution. Therefore, with FFT, the EEG signals represent a general state of consciousness over time which are hard to link to specific time-varying cognitive processes (Cohen, 2014).

Event Related De-synchronization (ERD) is an EEG method that circumvents the disadvantages of ERP and FFT by looking at both the time and frequency domain. Its strength is that it analyzes which frequencies occur, when they occur and how they change over time. Event-related synchronization (ERS) is a relative increase in power and ERD is a relative decrease. Alpha band de-synchronization has been observed during visual stimulation and also during cognitive and attentional tasks (Krause, 2000). Synchronization in the Theta band is also found, and this activity may be responsible for the encoding of new information (Klimesch, 1999; Klimesch, Doppelmayr, Russegger, & Pachinger, 1996). Because ERD covers a larger part of the EEG state space it could be a more sensitive and accurate measure than ERP and FFT. However, no ERD studies on vigilance and sustained attention have been reported.

The interest of the current study is to determine vigilance state during monotonous screening and monitoring tasks. During these tasks, critical signals are rare and most of the time consists of passively watching a screen. Staying alert therefore requires sustained attention over long periods of time, which is mentally demanding on the operator. Research on this topic is ongoing and the question is how to assess vigilance state during long periods of inactivity. To resemble a real working situation, a simple visual discrimination task is used consisting of mainly non-target events mixed with a few critical targets. Both stimuli occur irregular and participants need to respond to the targets only. Because participants are repeatedly exposed to the same visual non-target events, the focus will be on deriving vigilance state during periods when no critical signals are present. Vigilance state in this setting can only partly be derived from task performance and subjective scoring and therefore EEG will be used as a measure. It will be investigated, which analysis method yields the most usable EEG derived measure. In an exploratory analysis, the sensitivity of the ERP, FFT and ERD measures will be determined by looking at the predictive power on task performance, reaction time and subjective measure. Predictive power is derived by building separate regression models and select the best fitting model based on information criteria (Burnham & Anderson, 2002). ERD has not been used before with vigilance tasks. It is therefore expected that it is a more sensitive measure compared to ERP and FFT, because it can provide information from both the time and frequency domain.

Method

Participants

A total of nineteen participants took part in the experiment. All participants were students at the University of Twente and participated on a voluntary basis. The age of the group ranged from 20 to 29 with an average of $M = 24.2$ and $SD = 2.53$. The group consisted of 14 male, 5 female, 4 left-handed and 15 right-handed participants. All participants reported to have normal or corrected-to-normal vision and no neurological or psychiatric disorders. The study was approved by an ethical board of the University of Twente and all participants signed informed consent prior to the experiment.

Apparatus

The visual stimuli were presented on a Philips 17” CRT monitor, running at 60 Hz. with a screen resolution of 1024 x 768 pixels. A PC with 3.2 Ghz processor was used with Presentation software installed (Neurobehavioral systems, Inc.). All behavioral responses and EEG data were recorded with Brainvision Recorder software (version 1.05). EEG and EOG were amplified with a Quick-Amp amplifier (72 channels, DC), which implies an online average reference. Electrode impedance was kept below 10 kΩ. EEG and EOG were continuously recorded at a sampling rate of 1000 Hz. Online, a high cut-off filter was set at 200 Hz and a notch filter of 50 Hz was used.

Design

The focus of the study was on measuring EEG during prolonged periods of inactivity and compare the sensitivity of several EEG derived measures. No specific intervention or comparison between different conditions or groups was performed. The experimental manipulation consisted of using a simple monotonous visual monitoring task in a darkened environment. The independent variables were the mean amplitude peak and latency values for each EEG method and each block. The dependent variables were speed of response to the target stimuli, mean subjective state in each block and task performance in each block.

Task and procedure

All experiments were conducted during daytime. Participants were seated in an armchair behind the monitor at a distance of 60 cm. The lab room was completely darkened. The participants watched a short instruction about the experiment and started the experiment by pressing the space bar. Participants were instructed to look at a light grey central fixation cross in the middle of the screen, on a black background. The non-target stimuli consisted of two small light grey letters “M” that appeared simultaneously to the left and right from the central fixation. Sometimes a target stimulus appeared, which means that one of the letters “M” was exchanged with the letter “W”. In that case participants had to press the spacebar as fast as possible. No other specific criterion for detection was set and the letters were presented for 500 ms. The whole task lasted 40 minutes and consisted of 10 blocks of each 4 minutes. Each block consisted of 28 non-target trials, 4 target trials and 16 ‘empty’ trials containing no stimulus event. The duration of a trial was five seconds and an example of a sequence of events containing two consecutive trials (one non-target trial and one target trial) is displayed in Figure 1.

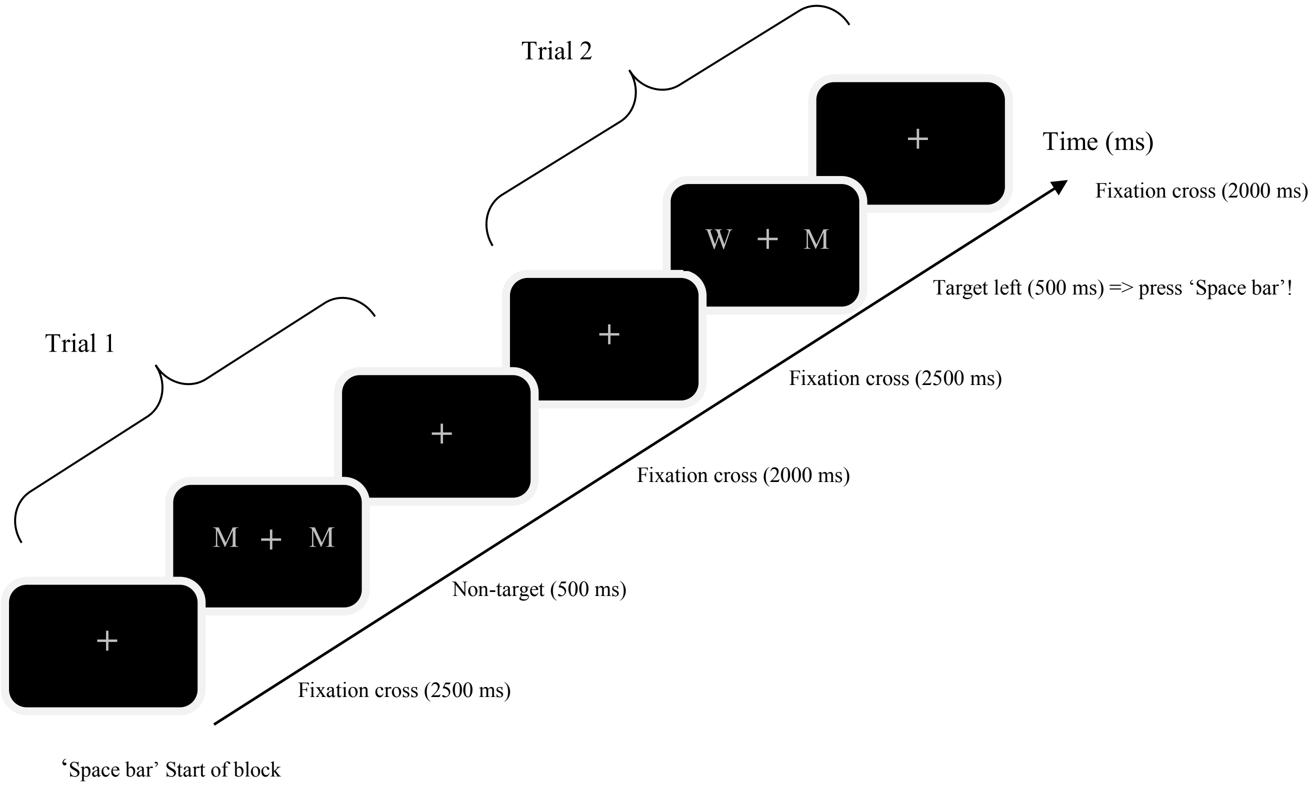

Figure 1. A schematic representation of two consecutive trials, containing a non-target event (Trial 1) and a target event (Trial 2).

During the experiment, both non-target and target trials appeared intermixed at a random rate. Targets could be located to the left or right and required a fast response by pressing the space-bar. By using randomization of all stimuli events, there was no influence of previous experience on the task and learning effects were minimal. All stimuli were equally divided to the left and right side to discourage saccades to the lateral locations. At the end of each block, each participant scored his or her level of alertness on a scale from 0 to 99 with paper and pencil. A new block was started by pressing the space bar.

Measures

During the task EEG was continuously recorded from 25 Ag/AgCl ring electrodes located at Fpz, F7, F3, Fz, F4, F8, FC5, FC6, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, PO7, PO3, PO4, PO8, and Oz. vEOG was recorded from electrodes placed above and below the left eye, and hEOG was measured from electrodes placed at the outer canthi of both eyes. From the EEG, different measures were offline derived by using ERP and ERD analysis on the non-target events and FFT analysis on non-event time-markers. A behavioral measure was derived from the speed of response to the target stimuli (reaction time in ms) and correct and incorrect detection and discrimination of both target- and non-target stimuli. The performance on the task was offline derived by classifying all correct and incorrect responses. All responses to the target stimuli were labelled as a ‘hit’ when a button was pressed and as a ‘miss’ when no button was pressed. All responses to the non-target stimuli were labelled as ‘correct rejection’ when no button was pressed or ‘false alarm’ when a button was pressed. Based on the number of classifications, the proportion correct rejections of the non-target stimuli (correct rejection rate) and the proportion of hits (hit-rate) for all target-stimuli were calculated for each block.

As a measure of performance, d-prime (d') was calculated. The sensitivity of this measure reflects the ability to discriminate between signal and noise, with a higher sensitivity leading to more hits and correct rejections and less false alarms and misses. Because the number of observations were too low to calculate a value for each block, one d-prime value was calculated over all blocks for every participant. To derive a performance measure at block level, a ‘Performance index’ was derived by summing the proportion of correct rejections and hits and subsequently dividing their sum by two. This is expressed by the following formula:

$$\frac{\left(\frac{n.o.\ correct\ rejections}{28}\right) + \left(\frac{n.o.\ hits}{4}\right)}{2}$$

In this way the Performance index is a measure of overall performance, in which the responses to both target and non-target events are equally weighted. The subjective measure consisted of a self-score on a scale from 0 (very drowsy) tot 99 (very alert). Participants were informed that a score of 50 would be considered as a normal state, 70 as an alert state and 30 as a drowsy state.

EEG data analysis

The raw EEG data were analyzed with Brain vision analyzer version 2.1 (Brain Products GmbH, 2014). To remove muscle and drift artifacts, a low cut-off filter of 0.1 Hz (24 dB/oct) was applied followed by a high cut-off filter of 35 Hz (24 dB/oct). Artifact rejection was set to semi-automatic with a gradient criteria of 100 µV/ms, minimum and maximum allowed voltage steps of +/- 250 µV and a low activity criterion of 0.1 µV with an interval length of 50 ms. A new baseline was set from -100 to 0 ms. An ocular correction for horizontal and vertical eye movements was applied with the Gratton and Coles algorithm (Gratton, Coles, & Donchin, 1983), followed by another artifact detection with minimal and maximal allowed amplitude set to -150 µV and 150 µV. A final check on artifacts was done in semi-automatic mode. In this way all individual segments were visually inspected on artifacts which were not detected in automatic mode. After that another baseline was set from -100 to 0 ms. To derive different measures from the EEG data, ERP, FFT and ERD analyses were conducted.

ERP analysis

The data were segmented based upon the non-target markers with a setting of -200 to 1000 ms. Averages for each person were calculated by using individual channel mode. In this mode, all electrode channels for each segment are checked separately for the presence of artifacts. For averaging, only the channels that contained artifacts were excluded while the other channels were still used.

Grand averages were made over all participants and for each block and these were visually displayed as transient views. A transient view shows the amount of EEG activity, where it is located and how its topography changes over time. The Grand averages revealed a P3b response which showed the highest amplitude value at electrode Pz (see Figure 2).

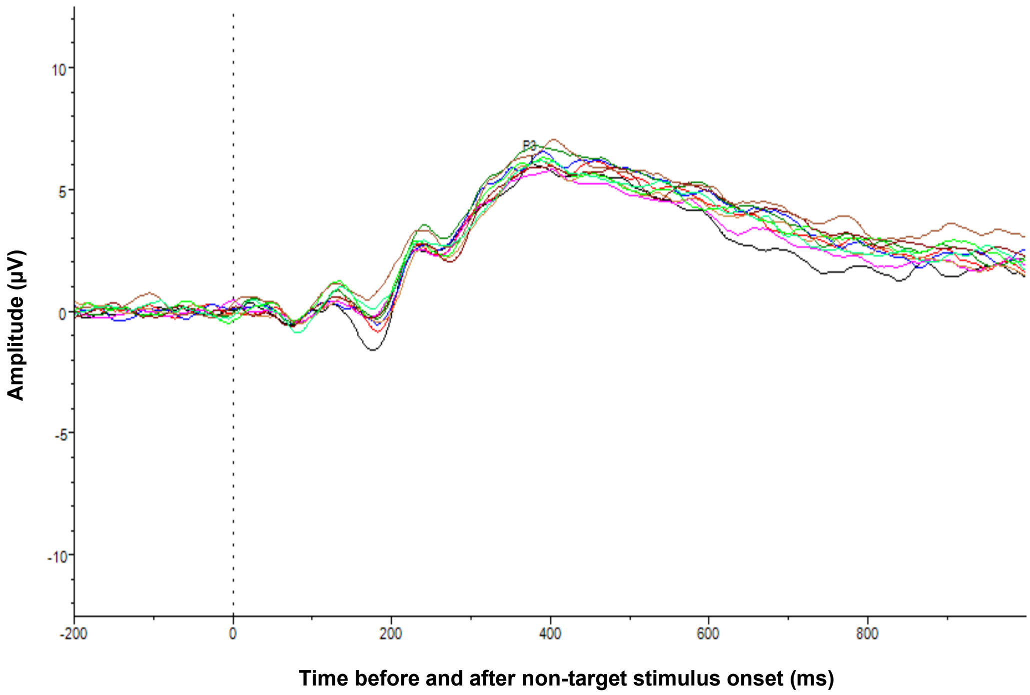

Figure 2. Grand average Event-related potential (ERP) P3b responses of electrode Pz for all blocks and over all participants.

Because an exploratory research was conducted with a new experimental task, electrode selection could not be based on previous evidence from literature and also the relations between the EEG and outcome variables were unknown. Therefore electrodes Fz, Cz, Pz and Oz were selected for further statistical analysis. For each testing person, the average peak and latency values of the P3b response were exported for each block. These were expressed in respectively µV and milliseconds.

FFT analysis

The EEG data were segmented based on the block-markers, with a setting from -200 ms to 3500 ms. A block-marker indicates the time-frame (block 1 to 10) of the recorded EEG signals and each block contained 48 of these markers, which were separated at least 3.5 seconds in time from the non-target and target stimuli. Average power values (μV^2) were calculated for every participant and each block by using individual channel mode. The average power values were calculated for the Delta band (1-4 Hz), the Theta band (4-8 Hz), the Alpha low band (8-10 Hz), the Alpha high band (10-12 Hz) and the Beta band (12-20 Hz).

Grand averages were made for every block and over all testing persons. Spectral views of the frequency distribution showed higher Theta band activity in frontal and central areas and higher Alpha activity (Alpha high) in occipital and parietal areas. Also individual Grand averages were made and its spectral and transient views were visually inspected and compared. They showed that for the Delta band, most activity occurred in the lower frequency range (1-1.5 Hz) at electrodes Fz, Cz and Pz. The Theta band showed most activity at electrodes Fz and Cz for almost all testing persons, which is displayed in Figure 3.

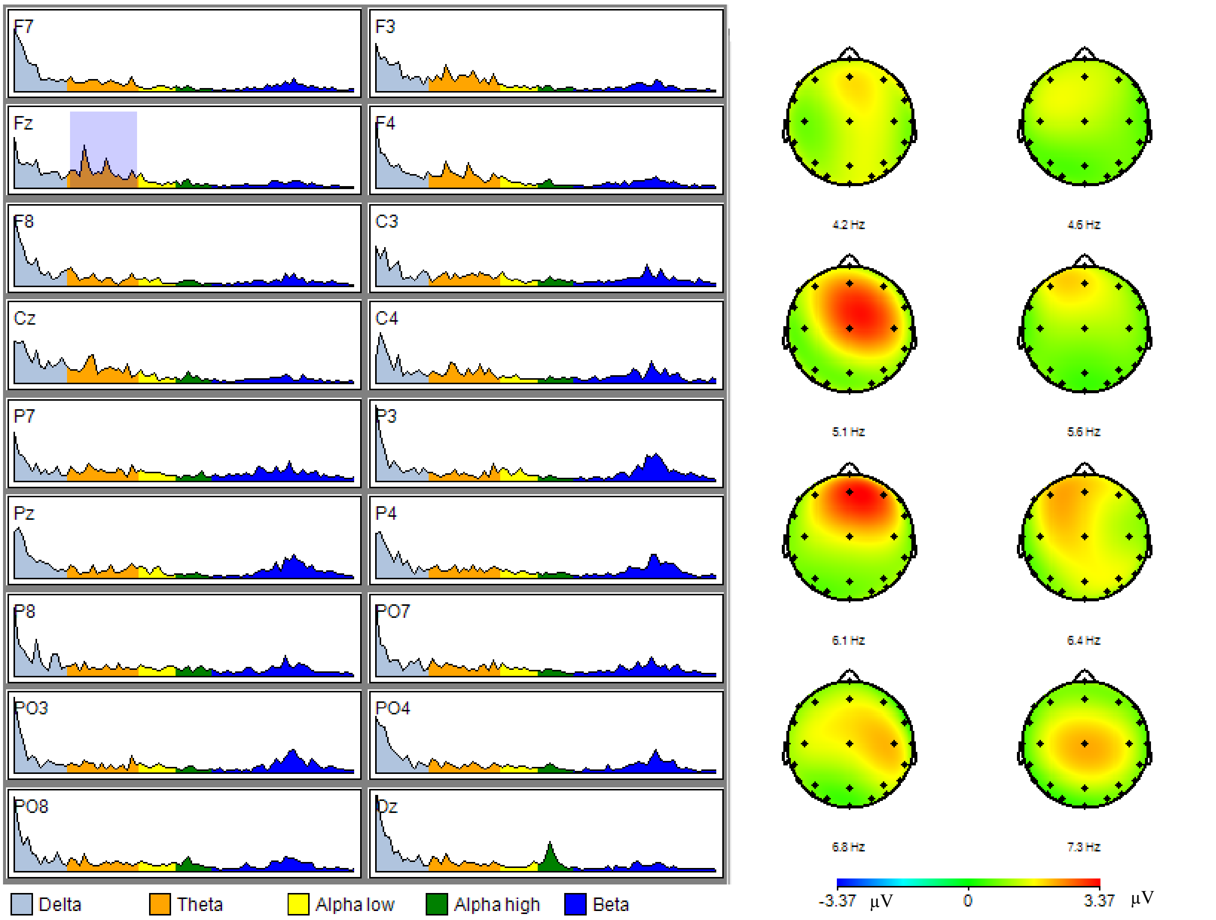

Figure 3. Example of Grand average Fast Fourier Transform (FFT) spectral views of 18 electrodes (left side) and a topographical view of the Theta band (right side) for one participant over all blocks.

For the Alpha low and Alpha high bands, most activity occurred at electrodes PO3 followed by PO4 and Pz, which is displayed in Figure 4.

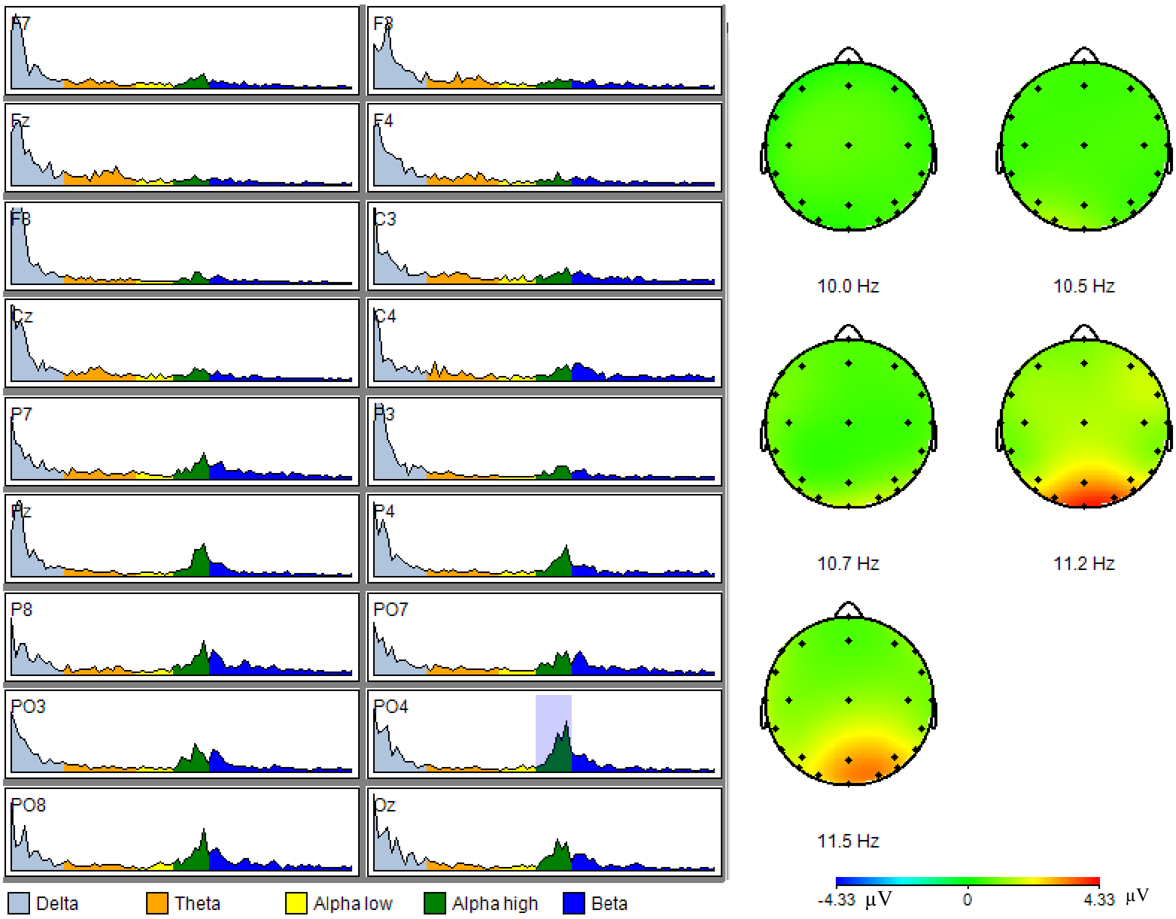

Figure 4. Example of Grand average Fast Fourier Transform (FFT) spectral views of 18 electrodes (left side) and a topographical view of the Alpha band (right side) for one participant over all blocks.

The Beta band showed most activity at 12 and 13 Hz located at electrode Pz for most participants.

Because a new experimental task was used for this study, all frequency bands were included for further statistical analysis, and the following electrodes were selected: for the Delta band electrodes Fz, Cz, Pz and Oz, for the Theta band electrodes Fz, Cz, F3 and F4, for both the Alpha low and Alpha high bands electrodes Pz, PO3, PO4 and Oz and for the Beta band electrodes Fz, Cz, Pz and Oz. For each testing person the average power values (μV^2) for every block were exported.

ERD analysis

The EEG data were segmented based upon the non-target markers with a setting from -1000 to 3500 ms. For every participant averages were calculated for each block by using individual channel mode. The average power values were calculated for the same frequency bands as used with the FFT analysis. The reference interval was set from -600 to -100 ms.

Grand averages were made for each block over all testing persons and for each frequency band. These were visually displayed as transient views, which showed the relative change in EEG activity, where this change was located and how its topography changed over time. They showed that the strongest synchronization occurred in the Delta band at electrode Cz, which is displayed in Figure 5.

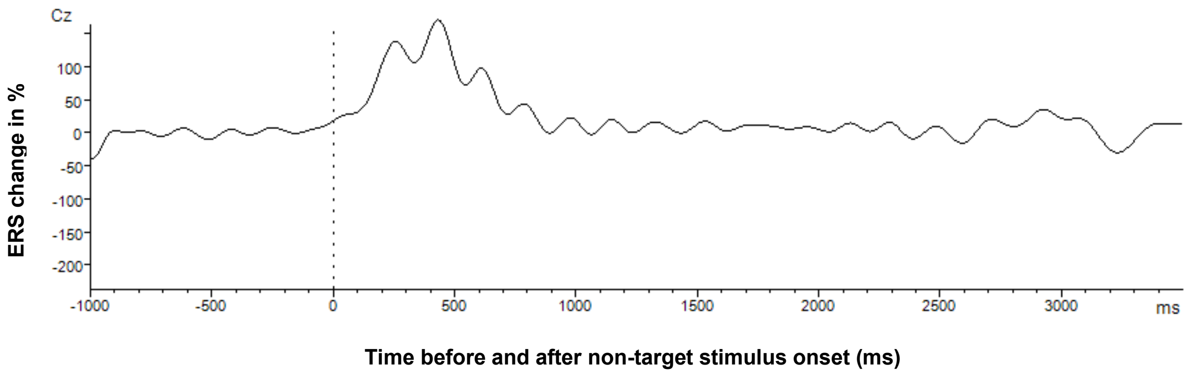

Figure 5. Grand Average Event-related synchronization (ERS) response over all participants and one block for the Delta band at electrode Cz.

The strongest de-synchronization occurred in the Alpha low and Alpha high bands at electrodes PO3 (see Figure 6) and PO4.

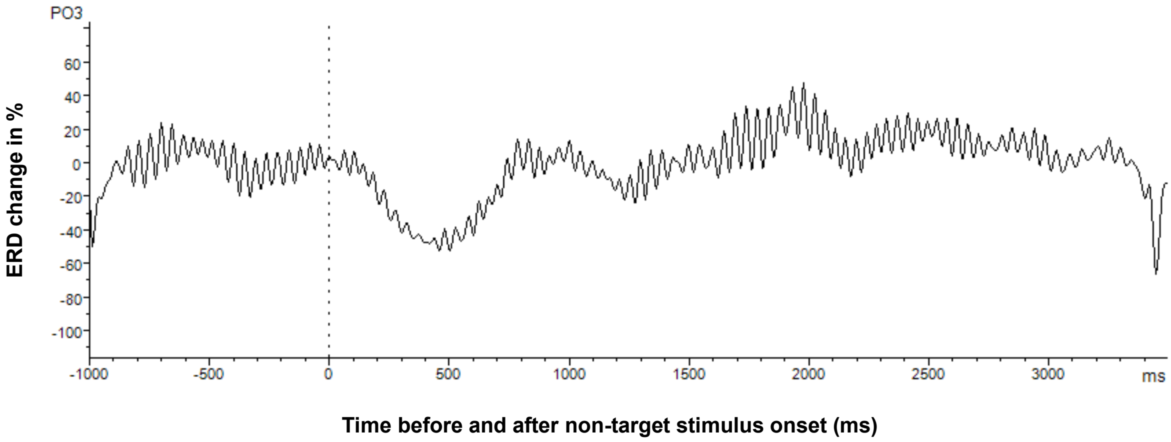

Figure 6. Grand Average Event-related desynchronization (ERD) response over all participants and one block for Alpha high band at electrode PO3.

For further statistical analyses, the following selection of electrodes was made; for both the Delta- and Theta bands, electrodes Fz and Cz and for both the Alpha low and Alpha high bands electrodes Pz, Oz, PO3 and PO4. For each testing person, the average amplitude peak (change in %) and latency values (ms) for every block were exported.

Data analysis

The purpose of the analysis was to compare the sensitivity of the EEG derived measures (ERP, FFT and ERD) regarding their predictive power on reaction time, the performance on the task and subjective state. To get a quick overview of general patterns in the data, descriptive and explorative analyses were conducted over all blocks. Box-plots and line-plots were made for all outcome variables as well as line-plots for the EEG predictor variables. For the performance on the task the absolute number of misses and false alarms for each block were calculated as well as the average number of hits, misses and false alarms per block.

After the descriptive and explorative analysis, separate regression analyses were conducted for each EEG derived predictor measure and each outcome variable. To account for dependency between measurements over all blocks, Generalized Estimating Equations (GEE) with gamma log link was chosen as the model type. GEE assumes observed data are correlated and therefore includes a dependence structure (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). The predictor variables were the P300/ERD peak and latency values and the average FFT power values of each block. The outcome variables were: mean subjective state, reaction time and task performance for each block.

Because no groups and conditions were compared and only measurements were conducted, hypothesis testing was not performed and p-values were not used. Basically, p-values only reveal if the tests show whether there is an association between variables, but statistical significance does not always indicate predictive value and comparing p-values is therefore not a measure of predictive power. Therefore, another approach was used in which the EEG derived measures were statistically compared by using information criteria, also called Akaike Information Criteria (AIC; Akaike, 1973). Information criteria are used as a means for model selection when deciding which variables need to be included in the regression (Pan, 2001). Given the data, a set of possible candidate models are compared to each other and AIC estimates the quality of each model by comparing them on how well they approximate reality. The preferred model is the one with the lowest AIC value because it minimizes information loss (the distance between reality and the approximating model) the most (Burnham & Anderson, 2002). In doing so, it deals with the trade-off between the goodness of fit of the model and the complexity of the model. AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of parameters that have to be estimated (Burnham & Anderson, 2002). Parsimonious models are preferred because they are less complex and contain fewer parameters to be estimated and therefore they involve less uncertainty in parameter estimation. Because GEE is non-likelihood based and takes repeated measurements into account, AIC cannot be directly applied. Instead a generalization of likelihood (a quasi-likelihood function) is used.

Regarding the current study, firstly regression models with all candidate predictor variables were built for every outcome variable. As a model structure, Auto-regressive correlation (AR1) was chosen. AR1 is observed when correlations between within-subject observations can be modelled directly as a function of the 'distance' between the observations in question (Zuur et al., 2009). Because the EEG was measured continuously over ten blocks, it was assumed that the measurements between directly neighboring blocks were strongly related and that the correlations would decrease for consecutive blocks that were located further away. To see how well the information in the data really fitted this correlation structure, the AR1 model was compared on Quasi-likelihood Information Criteria (QIC) values with those from independent, exchangeable and unstructured model structures. Independent structure assumes that the measurements are uncorrelated, exchangeable assumes homogenous correlations between elements and unstructured is a very general correlation model which, in essence, estimates all correlations between within-subject observations independently.

To determine which EEG predictors were relevant and should be included in the model, separate regression models were built for every EEG method and each EEG predictor. Based on the lowest Corrected Quasi-likelihood Information Criteria (QICC) values, the best EEG predictors were selected. It was found that adding more EEG predictors to the models did not improve the Corrected Quasi-likelihood Criteria (QICC) values. Adding more predictors is less parsimonious because more parameters have to be estimated. Therefore every model included only one predictor. From all 'one predictor models' the model with the best fit for every outcome variable was selected based on the QICC values. To determine which EEG analysis method was the best, The QICC values of all ‘best one-predictor models’ were compared.

Because a ‘best model’ is no assurance that it is actually a good predictor or explanation of the phenomenon of interest, all ‘best models’ were empirically validated to see how well they fitted the data and if they met the distribution assumptions of the GEE model. Therefore scatterplots were made in which the observed EEG predictor values were compared with the predicted values (XB predicted). Also QQ-plots (Q = quantile) of the raw residuals were made to test how well they fitted a normal distribution.

Results

Descriptive and explorative analyses of the EEG predictors and outcome variables

The boxplots for ‘Subjective state’ ($M = 55.54$, $SD = 1.22$), ‘Performance index’ ($M = 0.93$, $SD = 0.007$) and ‘Mean reaction time’ ($M = 607.037$, $SD = 7.32$) are listed in Appendix A1, A2 and A3. It shows that ‘Mean reaction time’ and ‘Performance index’ contain quite some outliers and ‘Subjective state’ tends to decline over time. This trend over time is clearly seen on the individual line-plots which are displayed in Appendix A4. The individual line-plots for ‘Mean reaction time’ and ‘Performance index’ did not reveal clear patterns over time and showed considerable variability.

Regarding the performance on the task, the absolute number of misses over all participants and all blocks was 61, with a mean per block of $M = 6.10$ and $SD = 3.035$. It comprised 8.03 percent of all target stimuli. The absolute number of false alarms was 299, which was higher than expected. The false alarms comprised 5.62 percent of all non-target stimuli and its mean per block was $M = 29.90$ and $SD = 9.723$. No participant missed all four targets in a block and one testing person missed all right targets. For each participant the average number of hits, misses and false alarms for each block were calculated. A histogram of the average number of hits (*M* = 3.68, *SD* = 0.66), misses (*M* = 0.32, *SD* = 0.66) and false alarms (*M* = 1.57, *SD* = 2.37) over all participants and all blocks is displayed in Figure 7.

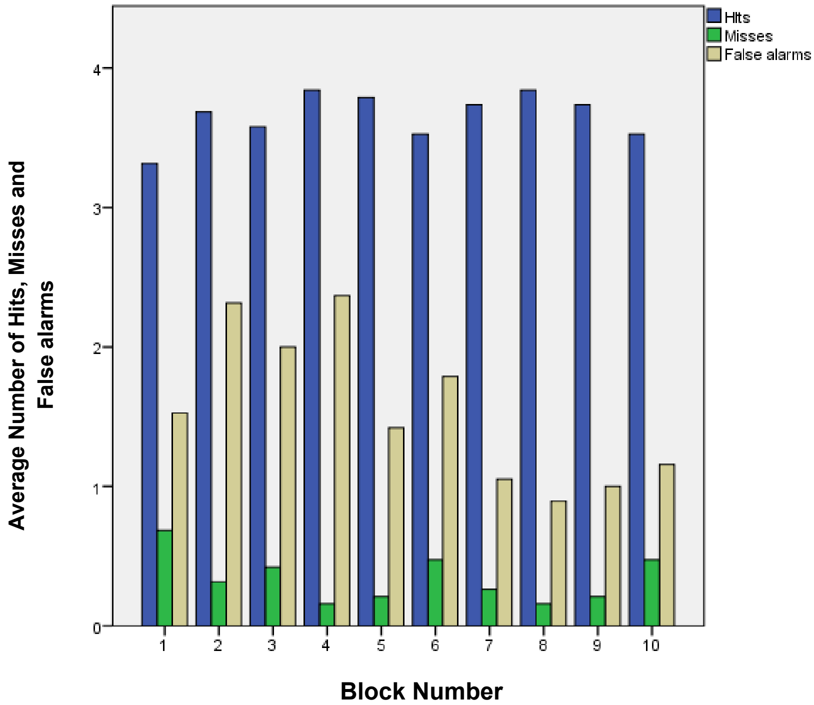

Figure 7. *Histogram of the average number of Hits, Misses and False alarms for each block.*

It shows that false alarms tend to decline over time. A total overview of the average number of hits, misses, correct rejections and false alarms per block is listed in Appendix A5. Based on this overview, d-prime (d') over all blocks was calculated for every participant. Because there were a lot of zero values for the misses, a correction was applied before calculating the hit- and false alarm rates. To the number of hits and false alarms, 0.5 was added and to the number of target stimuli and non-target stimuli 1 was added. D prime had a mean of *M* = 3.22 with *SD* = 0.696 and a range from 1.57 to 4.31.

The line-plots for the EEG predictor variables are listed in Appendix A6, A7, A8 and A9. The line-plots for FFT indicated a positive trend for Alpha low and Alpha high bands (see Appendix A6 and A7) on group level. For ERD (see Appendix A8 and A9) Delta and Theta bands showed a clear declining trend. Both trends were not found on the individual line-plots. For ERP, the line-plots did not reveal a trend over time.

Results ERP P3b

The QIC values for the best fitting correlation structures for each outcome variable and all candidate predictors are listed in Appendix B1. The Table shows that the AR1 model structure has the lowest QIC values for ‘Performance index’ (*QIC* = 23.269), ‘Reaction time’ (*QIC* = 36.307) and ‘Subjective state’ (*QIC* = 37.768). This indicates that the data fit this model structure the best. For ‘Performance index’ the difference between Independent (*QIC* = 23.604) and AR1 is small.

The QICC values for all separate ‘one-predictor models’ are displayed in Appendix B2. The Table shows that for both ‘Reaction time’ and ‘Subjective state’ the model based on electrode Oz amplitude contains the best predictor with QICC values of 8.614 and 24.784. For ‘Reaction time’ the differences with the model based on electrode Fz amplitude (*QICC* = 8.618) and the model based on electrode Pz amplitude (*QICC* = 8.657) are small. For ‘Performance index’ the model based on electrode Fz latency proves to have the best predictor with a QICC value of 6.543, but the difference with the other models are very small.

Results FFT

The QIC values for the best fitting correlation structures for each outcome variable and all candidate predictors are listed in Appendix B3. For ‘Reaction time’ and ‘Subjective state’ AR1 proves to be the best model structure with the lowest QIC values of respectively 36.307 and 39.848. For ‘Performance index’ the Exchangeable structure has a slightly better fit (*QIC* = 36.314) compared to AR1 (*QIC* = 36.453). However, the Exchangeable structure assumes that correlations between all blocks are homogenous, which is not very likely. Therefore AR1 was chosen as the preferred model structure. For AR1 the adjacent blocks showed a correlation of r = 0.304, which declined for secondly adjacent blocks (r = 0.093) and thirdly adjacent blocks (r = 0.028).

The QICC values for all separate ‘one-predictor models’ are listed in Appendix B4. The Table shows that the model based on Alpha low band at electrode Pz has the lowest QICC value (*QICC* = 6.399) for ‘Performance index’. For ‘Reaction time’ the model based on Alpha high band at electrode Pz has the best predictor with a QICC value of 8.437 and for ‘Subjective state’ the model based on Beta band at electrode Pz has the best predictor with a QICC value of 25.343.

Results ERD

The QIC values of the correlation structures for each outcome variable and all candidate predictors are listed in Appendix B5. The table shows that for ‘Subjective state’ AR1 is the best model structure with a QIC value of 40.500. For ‘Performance index’ and ‘Reaction time’ ‘Unstructured’ indicates a better model fit. Its QIC values are respectively 39.195 and 39.990 while AR1 QIC values are respectively 41.870 and 44.757. However, ‘Unstructured’ doesn’t model observations directly as a function of the distance between observations and needs a lot more parameters to be estimated. Therefore AR1 was chosen as the preferred model structure.

The QICC values for all separate ‘one-predictor models’ are listed in Appendix B6. The Table shows that for ‘Performance index’ and ‘Reaction time’ the model based on Alpha low band amplitude at electrode Pz has the best predictors with QICC values of respectively

6.458 and 8.662. For Subjective state, the model based on Delta band amplitude at electrode Cz has the best predictor with a QICC value of 25.097.

To determine which EEG method yielded the most predictive EEG measure, the best model from each analysis method was selected and all best models were compared on QICC values for each outcome variable. The QICC values are listed in Table 1.

Table 1

GEE QICC values of the best one-predictor EEG models for Performance index, Reaction time and Subjective state.

Outcome variable	ERP			FFT			ERD		
	Band	Electrode	QICC	Band	Electrode	QICC	Band	Electrode	QICC
Perf. index	-	Lat. Fz	6.543	Alpha low	Pz	6.399	Alpha low	Pz	6.458
Reaction time	-	Oz	8.614	Alpha high	Pz	8.437	Alpha low	Pz	8.662
Subjective state	-	Oz	24.784	Beta	Pz	25.343	Delta	Cz	25.097

Notes: GEE = Generalized Estimating Equations, QICC = Corrected Quasi-likelihood Information Criteria, EEG = Electro-encephalography, Perf.index = Performance Index, Lat. = Latency, ERP = Event-related potential, FFT = Fast Fourier Transform, ERD = Event-related desynchronization. Lower QICC values are better due to less information loss.

According to the information criteria, FFT yielded the best predictive models for Performance index and Reaction time, but the differences with ERP and ERD were relatively small. For Subjective state, ERP yielded the best predictive model.

To see how good the selected 'best models' fitted the data, they were empirically validated. Each separate model was run and the regression coefficients are listed in Table 2.

Table 2

GEE Regression coefficients for 'best EEG models' on Performance index, Reaction time and Subjective state.

Best model	Outcome variable	Predictor		B	SE *(B)*	Sig. *(p)*
		Band	Electrode			
FFT	Performance index	Alpha low	Pz	-0.018	0.0069	0.009*
FFT	Reaction time	Alpha high	Pz	.016	.0115	0.177
ERP	Subjective state	-	Oz	0.006	0.0023	0.007*

Notes: * = significant at $p < 0.05$. GEE = Generalized Estimating Equations, EEG = Electro-encephalography, FFT = Fast Fourier Transform, ERP = Event-related potential.

For 'Performance index' FFT Alpha low Pz proved indeed to be predictive. It shows a negative relation with $B = -0.018$ and $p = 0.009$. This means that a one unit increase of the Alpha low Pz predictor is associated with a 0.018 decrease in Performance index score. Also, for 'Subjective state' ERP electrode Oz proved to be predictive. It shows a positive relation with $B = 0.006$, and $p = 0.007$. This means that a one unit increase of the ERP electrode Oz predictor is associated with a 0.006 increase in Subjective state score. For 'Reaction time' FFT Alpha high PZ did not prove to be predictive ($p > 0.05$).

For all three models, QQ-plots of the residuals were made as well as scatter-plots of the model predicted and observed values (see Appendix C). It shows that for Performance index (see Appendix C1) and Reaction time (see Appendix C3) the residuals do not quite fit a normal distribution, which partly violates the assumption of GEE. Only the residuals for Subjective state (see Appendix C5) fit a normal distribution well. The scatterplots show that for Performance index (see Appendix C2) and Mean reaction time (see Appendix C4), the predicted and observed values do not show a clear relation. This indicates that these models are not very accurate, because a considerable amount of the residuals are left unexplained. For Subjective state (see Appendix C6), the predicted and observed values show a clear relation, so the model can explain a considerable amount of the residuals.

Discussion

EEG is considered a promising vigilance measure and the current study explored which analysis method yielded the most sensitive EEG derived measure. A monotonous vigilance task was used with few and irregular stimuli events. It was found that FFT proved to be the most informative analysis method for ‘Performance index’ and ‘Reaction time’ and for ‘Subjective state’ ERP proved to be most informative. Therefore the results did not confirm the main hypothesis which stated that ERD would yield the most sensitive measure.

The results showed that the differences between the three analysis methods were very small, so it is not very clear which method would be the most useful. It was also observed that for each analysis method the information criteria of its predictor-models did not much differ on the outcome variables. Therefore caution should be considered in drawing conclusions about the predictive power and in generalizing the findings.

For FFT, Alpha band at parietal site proved to be the most informative predictor for both ‘Performance index’ and ‘Reaction time’. In literature, ‘Reaction time’ is considered to be strongly related to ERP P3 latency. P3 latency increases with the time required to detect and evaluate a target stimulus (Polich, 2007). Although ERP P3 latency did not prove to be the best predictor, the differences between FFT and ERP method were rather small and the model validation showed that the FFT model was actually not a good predictor for reaction time.

FFT Alpha band showed an increase over time which could be indicative of mental inactivity due to the monotony of the task (Molina, Correa, Sanabria, & Jung, 2013). Contrary to what might be expected, Performance on the task did not decrease over time. It seemed that during this task sustained attention could be maintained over a prolonged period without much effort. Frontal Theta band did not show an increase over time, which indicates that the task was mentally low-demanding on cognitive resources and vigilance state remained constant over time. It is possible that participants used a form of self-regulation strategy to cope with the monotonous nature of the task and keep performance at the same level. It must also be noted that some participants were taking regular breaks between every block, which could be used to prevent resources from getting drained. It was observed that for some persons these breaks lasted more than 5 minutes.

Although performance on the task remained constant, all participants reported an increase in drowsiness and decrease in alertness over time. It can be questioned how the subjective scores should be interpreted. Perceived drowsiness and alertness are linked to the availability of resources and mental effort to the task (Matthews & Davies, 2001; Smit, 2004), but it is more likely that the scores represented a state of under-arousal caused by physical inactivity during the task and also induced by the darkened room. It could also be indicative of declining motivation and increasing boredom which could disengage participants from the task.

A remarkable finding was that participants performed very well on the task, better than would be expected based on literature. Firstly, in general humans are not well suited for the detection of low probability events over longer periods (Warm, Dember, & Hancock, 1996). Secondly, the task should be harder to perform since it contained fewer stimuli events and the non-target stimuli occurred irregular. The results showed a high hit-rate with many zero scores for the misses. This indicates that the task used was too easy, which would make the performance measure not very discriminative. A possible explanation is that, due to the low amount of stimuli, the target stimuli could be more easily detected and the infrequent appearance of both target and non-target stimuli braked monotony and kept participants alert.

The performance index might not be the most optimal measure and its results are probably biased. It is hard to tell if misses were caused by lapses of attention or by individual differences in ability to concentrate on the task. Also, from the correct rejections it can't be derived if it represents discrimination between stimuli events or if the non-target stimuli were detected at all. Because the false alarm rate was higher than expected, it could be that some participants pressed more easily on the space bar, to raise the chances for a hit. The false alarms also showed large variability between participants, which probably biased the results. In general, the false alarm rates were low, but for some participants the false alarm rate was very high.

The recorded EEG signals were not very clean which might have influenced the results. Although ERD is considered a sensitive measure, it is also much more sensitive to noise because all the data are considered as a signal. This could easily lead to a distortion in the ERD baselines used. Compared to ERD, ERP and FFT are less sensitive to noise. With ERP most noise gets lost through the averaging procedure and FFT doesn't use a baseline. Next to the amount of noise, the EEG predictor variables but also the performance variables showed large inter- and intra-variability. This makes it hard to infer coherence between the different measures on group level.

The large amount of variability in the predictor and outcome variables can possibly explain why the three best selected EEG models did not fit the data very well. The model validation showed that for Performance index and Reaction time a considerable amount of the residuals could not be explained. This implicates that they did not properly separate noise from structural information and that the assumptions of GEE were not fully met. Only

Subjective state showed a good model fit with the data. Although it showed the least variability, the EEG predictors were less informative on predicting this outcome variable.

A limitation of the study was that using one-predictor EEG models might be an oversimplification of reality. Although parsimonious models are easy to interpret and have a more precise parameter estimate, they could also be too simple. As the analyses have shown, EEG is not just a simple metric because these signals are very complex. Vigilance is probably reflected in multiple frequency bands (Berka et al., 2007) and therefore multiple classifiers need to be derived from a complex combination of EEG variables (Blankertz et al., 2010; Finomore et al., 2009; Müller, Tangermann, Dornhege, Krauledat, Curio, & Blankertz, 2008) in order to extract predictive information.

Another limitation of this study was that no pre- and post-test measurements were performed. Because there was no baseline reference on individual level, the EEG signals could not be compared before, during and after the task, which makes it hard to tell what the influence of the vigilance task was. It was also hard to compare the results of this study with findings from similar studies. Because a new experimental task was used, the manipulations were not exactly the same which made it unclear what the current findings actually represented and how they should be interpreted.

For future research it is suggested to make the vigilance task sufficiently discriminable between hits and misses and to slightly increase the workload. Therefore, the current task should be extended to approximately two hours, a pre-test should be included and the breaks between the blocks should be removed. Regarding the stimuli used, the target stimuli should be kept unpredictable, but there should be enough space between them and they should not appear directly at the beginning of a block. Because ERD might not be the most optimal and informative EEG method, it should be considered to use or combine it with Wavelet- analysis. Wavelets also take both the time and frequency-domain into account but do not use a baseline, so changes in the EEG are displayed in absolute instead of relative values.

To conclude this discussion, the current study showed that EEG signals are predictive of behavioral and subjective vigilance measures, but the differences found were rather small. The results on the vigilance task used need to be compared with similar studies to understand what the found brain-signals actually represent and how they should be interpreted. For practical applications, future research needs to take individual differences into account and specify in detail which relevant information of the EEG signals needs to be derived.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B.N. Petrov (Ed.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., … Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space and Environmental Medicine*, *78*(5).

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., …Müller, K.R. (2010). The Berlin brain-computer interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience*, *4(*198*)*.

Burnham, K.P., & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York, NY: Springer.

Cohen, M. (2014). *Analyzing Neural Time Series Data: Theory and Practice.* Cambridge, MA: MIT Press.

Eichele, H., Juvodden, H.T., Ullsperger, M., & Eichele, T. (2010). Mal-adaptation of event-related EEG responses preceding performance errors. *Frontiers in Human Neuroscience*, *10*(4).

Finomore, V., Warm, G., Matthews, T., & Shaw, J. (2009). Predicting vigilance: A fresh look at an old problem. *Ergonomics*, *52*(7), 791-808.

Gevins, A., & Smith, M. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 113-131.

Gratton, G., Coles, M.G.H., & Donchin, E. (1983). A new method for off-line removal of ocular artifacts. *Electroencephalography and Clinical Neurophysiology, 55*, 468–484.

Hancock, P.A. (2013). In search of vigilance: the problem of iatrogenically created psychological phenomena. *American Psychologist, 68*, 97–109.

Hopfinger, J.B., & West, V.M. (2006). Interactions between endogenous and exogenous attention on cortical visual processing. *NeuroImage*, *31*(2), 774–789.

Kamzanova, A.T., Kustubayeva, A.M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *56*(6), 1136-1149.

Klimesch, W., Doppelmayr, M., Russegger, H., & Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. *NeuroReport, 7*, 1235–1240.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2-3), 169-195.

Krause, C.M., (2000). Auditorily Elicited Event-related Desynchronization (ERD) and Synchronization (ERS) as a Method for Studying Cortical Correlates of Cognitive Processes. *Revista Espanola de Neuropsicologia*, *2*(1-2), 77-91.

Langner, R., & Eickhoff, S.B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139*(4), 870-900.

Lim, J., & Dinges, D.F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*, 305–322.

Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*, 6–21.

Martel, A., Dähne, S., & Blankertz, B. (2014). EEG predictors of covert vigilant attention. *Journal of Neural Engineering*, *11*(3), 11.

Matthews, G., & Davies, D.R. (2001). Individual differences in energetic arousal and sustained attention: A dualtask study. *Personality and Individual Differences*, *31*, 575-589.

Molina, E., Correa, A., Sanabria, D., & Jung, T.-P. (2013). Tonic EEG Dynamics During Psychomotor Vigilance Task. *Proceedings of the 6th Annual International IEEE EMBS Conference on Neural Engineering* (pp. 1382-1385). doi:10.1109/NER.2013.6696200

Müller, K.R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods, 167*(1), 82-90.

Nickerson, R.S. (1992). *Looking ahead: Human factors challenges in a changing world*. Mahwah, NJ: Erlbaum.

O' Connell, R.G., Dockree, P.M., Robertson, I.H., Bellgrove, M.A., Foxe, J.J., & Kelly, S.P. (2009). Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal Of Neuroscience, 29*(26), 8604-8611.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, 57(1), 120-125.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-48.

Robertson, I.H., & Garavan, H. (2004). Vigilant attention. In: M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences, 3rd edition*. (pp. 563-578). Cambridge, MA: MIT Press.

Rosvold, H., Mirsky, A.F., Sarason, I., Bransome, E.D., & Beck, L.H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20*, 343–350.

Smit, A.S. (2004). *Vigilance or availability of processing resources: a study on cognitive*

Energetics (Doctoral dissertation). Retrieved from http://repository.ubn.ru.nl/handle/2066/19548

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433-441.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman (Ed.), *Automation and human performance: Theory and applications* (pp. 183-200). Hillsdale, NJ: Erlbaum.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: a Psychophysiological analysis of the reciprocity of information-processing resources. *Science 9*(221), 1080–1082.

Zordan, L., Sarlo, M., & Stablum, F. (2008). ERP components activated by the GO! And WITHHOLD! conflict in the random sustained attention to response task. *Brain and Cognition 66*, 57-64.

Zuur, A.F., Ieno, E.N. Walker, N.J., Saveliev, A.A., & Smith, G.M. (2009). Generalised Estimation Equations. *Mixed effects models and extensions in ecology with R* (pp. 295-321). New York, NY: Springer.

Appendix A

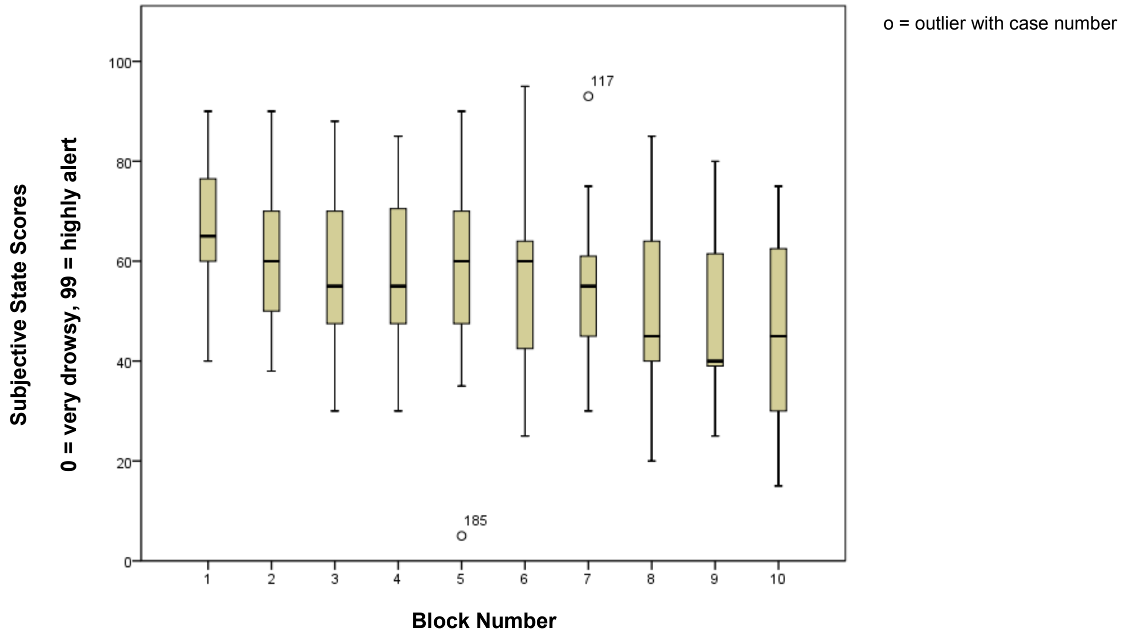

Figure A1 Boxplots of Subjective state.

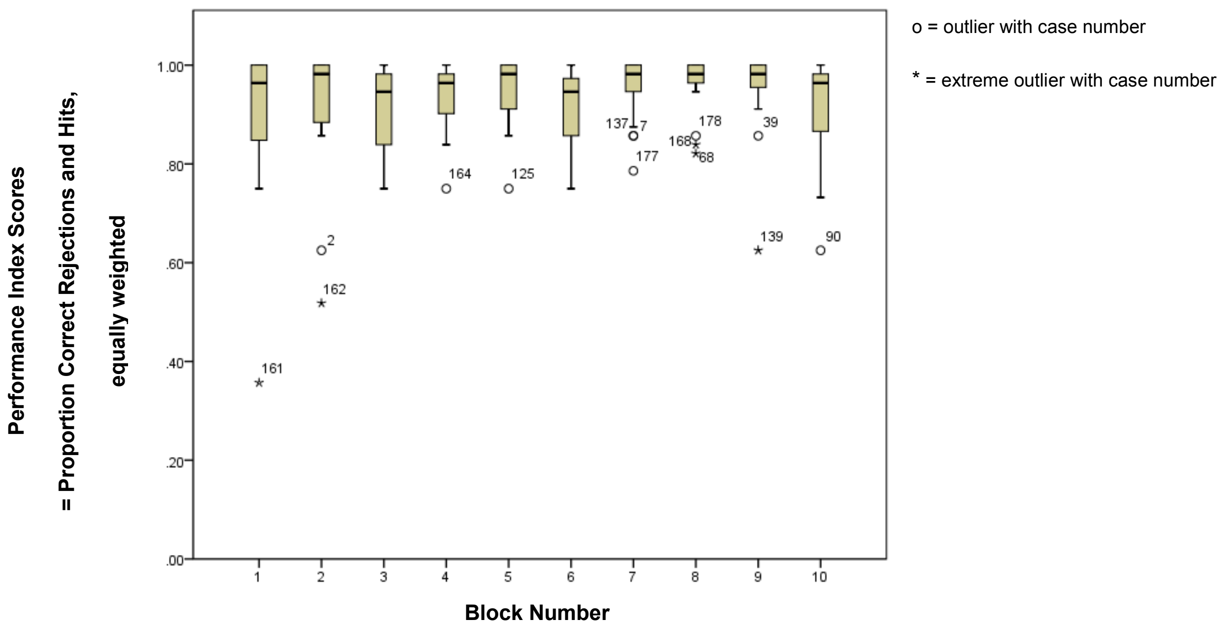

Figure A2 Boxplots of Performance index.

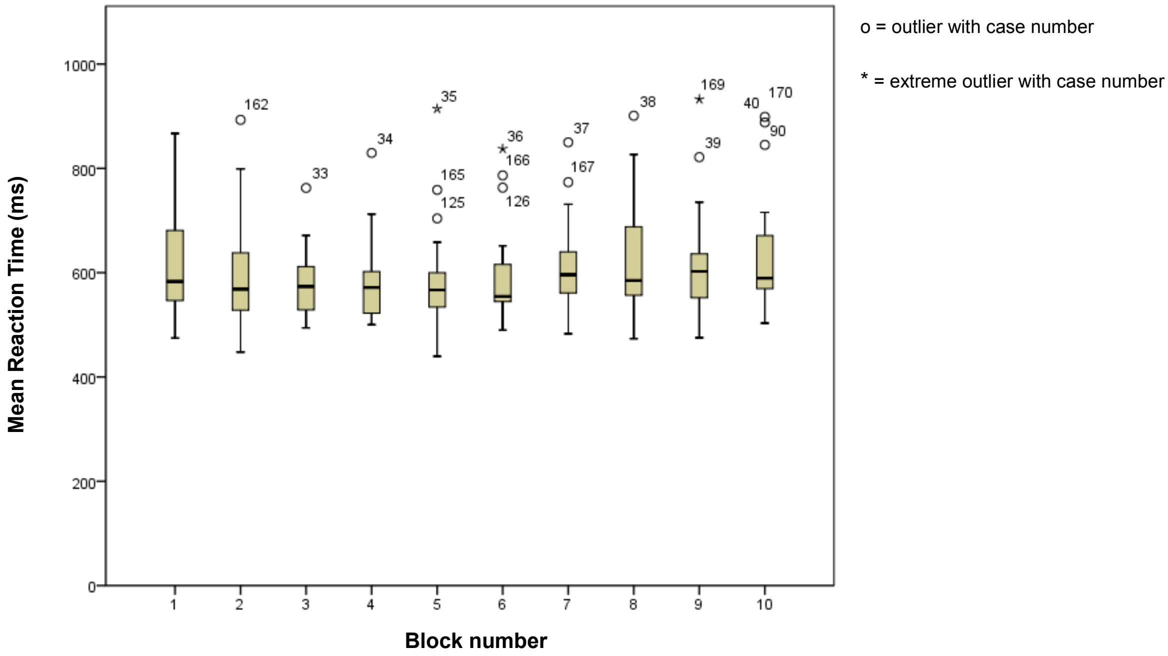

Figure A3 Boxplots of Mean reaction time.

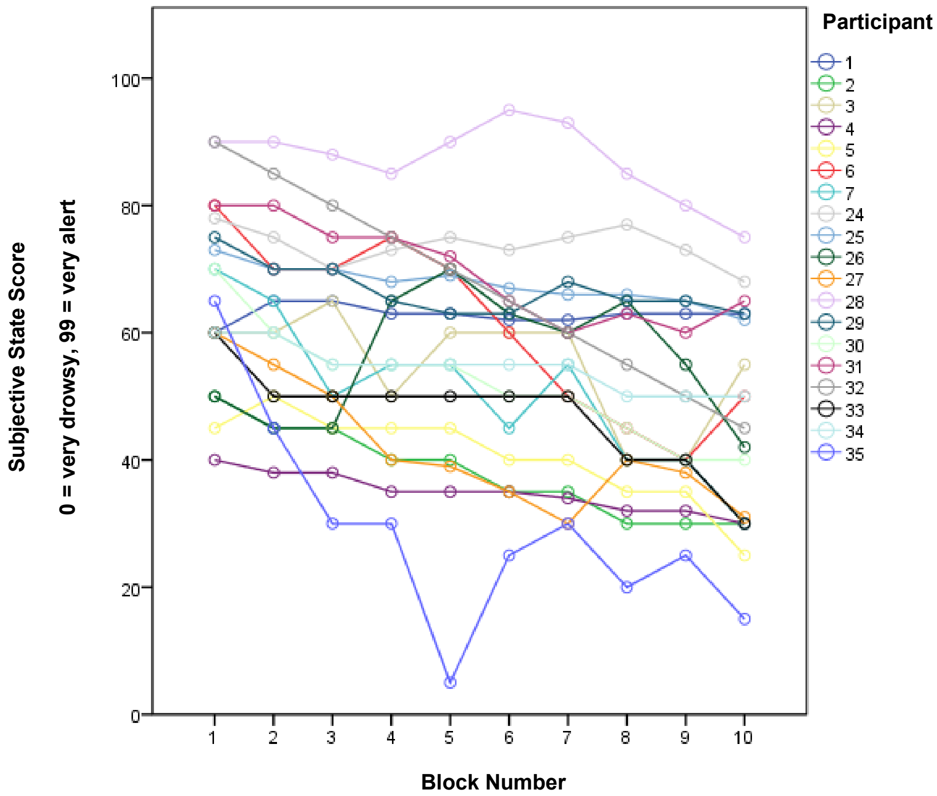

Figure A4 Individual line-plots of Subjective state scores over all blocks.

Table A5

Average number of hits, misses, correct rejections and false alarms for each block.

Block	1	2	3	4	5	6	7	8	9	10
Hits	3.32	3.68	3.58	3.84	3.79	3.53	3.74	3.84	3.74	3.53
Hitrate	0.842	0.921	0.882	0.961	0.947	0.895	0.947	0.934	0.947	0.855
Misses	0.68	0.32	0.42	0.16	0.21	0.47	0.26	0.16	0.21	0.47
CR	26.47	25.68	26.05	25.63	26.58	26.21	26.89	27.05	26.89	26.79
CR rate	0.946	0.917	0.930	0.915	0.949	0.936	0.960	0.966	0.960	0.957
FA	1.53	2.32	2.00	2.37	1.42	1.79	1.05	0.89	1.00	1.16

Notes: CR = Correct Rejections, FA = False alarms. Each block contained 4 target events and 28 non-target events. The target events were classified as hits and misses and the non-target events as correct rejections and false alarms. A Hits-score of 3.32 means that, on average, participants classified more than 3 out of 4 target events correctly. Hitrates and CR rates were calculated by dividing the individual Hit- and CR-scores by respectively the total number of targets and non-targets in each block.

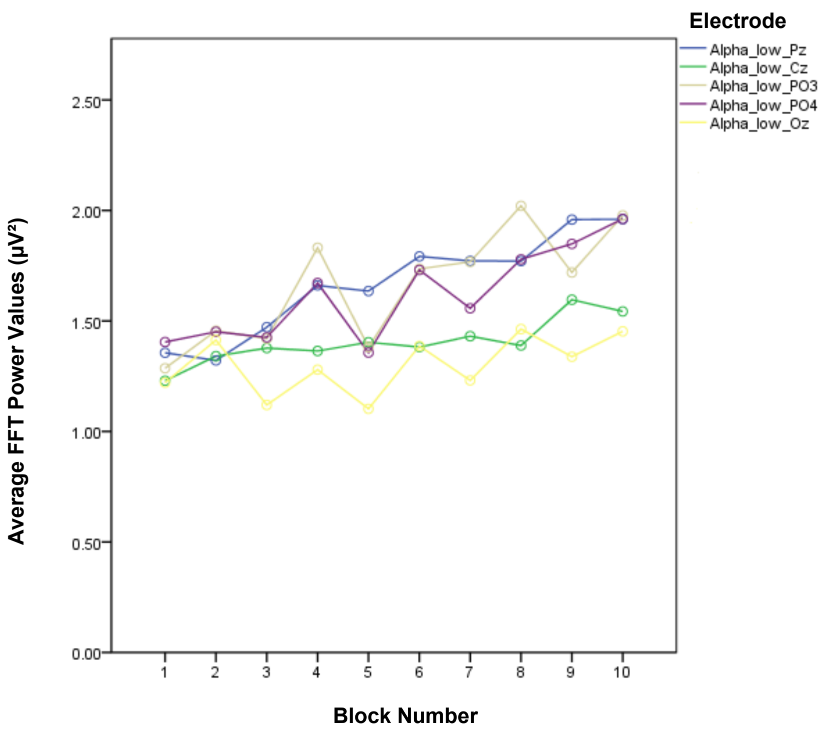

Figure A6 Line-plots of group average FFT power values for Alpha low band.

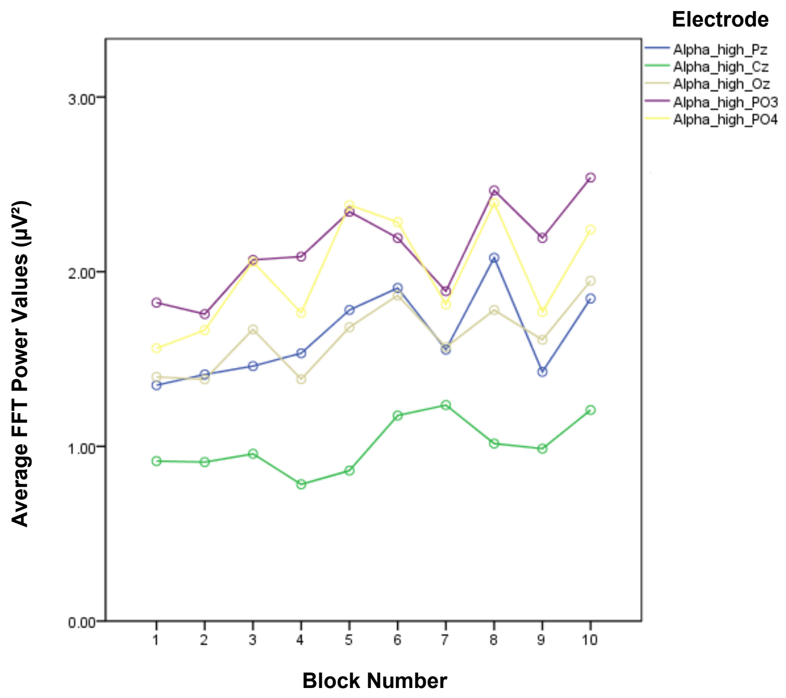

Figure A7 Line-plots for group average FFT power values for Alpha high band.

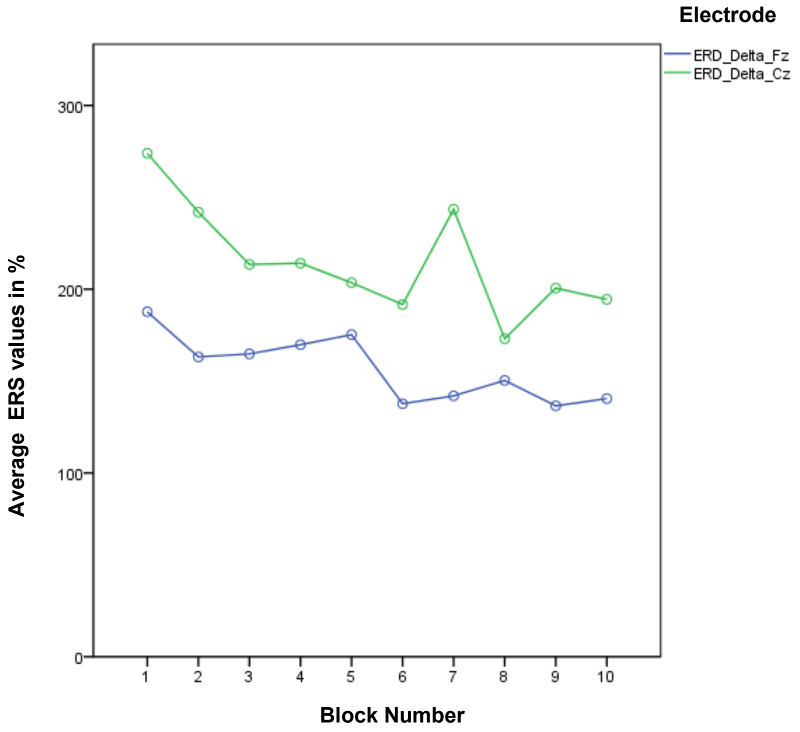

Figure A8 Line-plots for group average ERD peak values for Delta band.

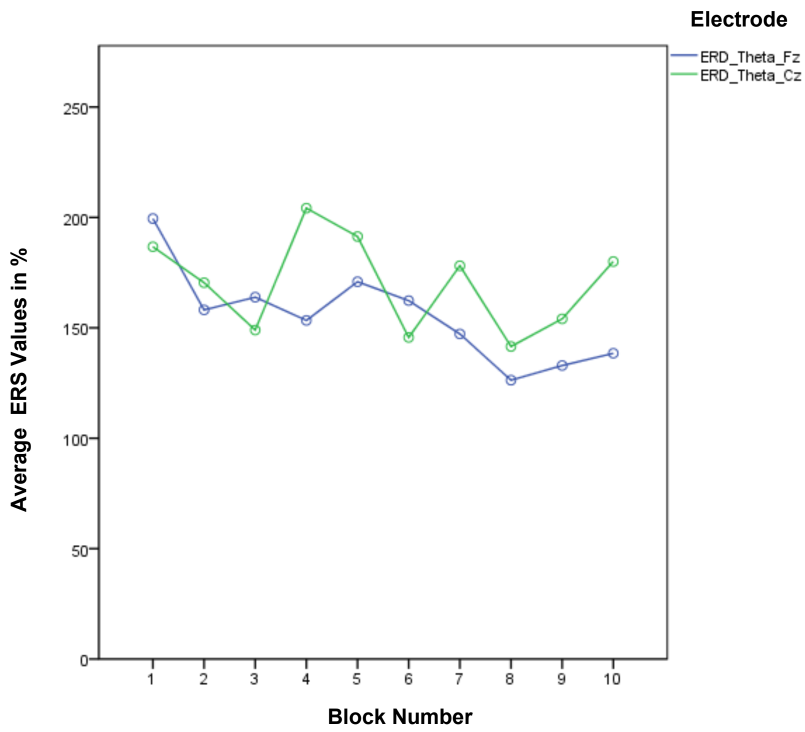

Figure A9 Line-plots for group average ERD peak values for Theta band.

Appendix B

Table B1

GEE QIC values for Independent, AR1, Exchangeable and Unstructured correlation structures for each outcome variable and all candidate ERP predictors.

Correlation structure	Outcome variable		
	Performance index	Reaction time	Subjective state
Independent	23.604	71.971	80.872
AR1	23.269	36.307	37.768
Exchangeable	39.038	38.011	49.222
Unstructured	49.282	194.729	78.327

Note: GEE = Generalized Estimating Equations, QIC = Quasi-likelihood Information Criteria, AR1 = Auto-regressive, ERP = Event-related potential. Lower QIC values are better due to less information loss.

Table B2

GEE QICC values of all one-predictor ERP regression models for all outcome variables.

ERP regression model	Outcome variable		
	Performance index	Reaction time	Subjective state
Model based on electrode Fz amplitude	6.548	8.618	25.386
Model based on electrode Cz amplitude	6.544	8.753	25.442
Model based on electrode Pz amplitude	6.548	8.657	25.822
Model based on electrode Oz amplitude	6.545	8.614	24.784
Model based on electrode Fz latency	6.543	8.669	25.555
Model based on electrode Cz latency	6.548	8.778	25.095
Model based on electrode Pz latency	6.547	8.754	25.318
Model based on electrode Oz latency	6.555	8.781	25.461

Notes: GEE = Generalized Estimating Equations, QICC = Corrected Quasi-likelihood Information Criteria, ERP = Event-related potential. Lower QICC values are better because due to less information loss.

Table B3

GEE QIC values for Independent, AR1, Exchangeable and Unstructured correlation structures for all candidate FFT predictors and each outcome variable.

	Outcome variable		
Correlation structure	Performance index	Reaction time	Subjective state
Independent	36.619	71.971	96.024
AR1	36.453	36.307	39.848
Exchangeable	36.314	38.011	58.402
Unstructured	86.335	194.729	41.941

Notes: GEE = Generalized Estimating Equations, QIC = Quasi-likelihood Information Criteria, AR1 = Auto-regressive, FFT = Fast Fourier Transform. Lower QIC values are better due to less information loss.

Table B4

GEE QICC values of all one-predictor FFT regression models for all outcome variables.

	Outcome variable		
FFT regression model	Perform. index	RT	Subjective state
Model based on Delta band electrode Fz power	6.513	8.752	25.459
Model based on Delta band electrode Cz power	6.482	8.760	25.456
Model based on Delta band electrode Pz power	6.513	8.728	25.433
Model based on Delta band electrode Oz power	6.483	8.761	25.595
Model based on Theta band electrode Fz power	6.551	8.758	25.463
Model based on Theta band electrode Cz power	6.551	8.748	25.399
Model based on Theta band electrode F3 power	6.548	8.816	25.518
Model based on Theta band electrode F4 power	6.532	8.761	25.470
Model based on Alpha low band electrode Pz power	6.399	8.536	25.447
Model based on Alpha low band electrode PO3 power	6.461	8.509	25.601
Model based on Alpha low band electrode PO4 power	6.471	8.587	25.617
Model based on Alpha low band electrode Oz power	6.515	8.526	25.567

(Continued)

Table B4 (Continued)

FFT regression model	Outcome variable: Perform. index	RT	Subjective state
Model based on Alpha high band electrode Pz power	6.513	8.437	25.443
Model based on Alpha high band electrode PO4 power	6.503	8.459	25.476
Model based on Alpha high band electrode Oz power	6.546	8.675	25.460
Model based on Beta band electrode Fz power	6.519	8.705	25.525
Model based on Beta band electrode Fz power	6.532	8.560	25.470
Model based on Beta band electrode Pz power	6.518	8.877	25.343
Model based on Beta band electrode Oz power	6.549	8.710	25.464

Notes: GEE = Generalized Estimating Equations, QICC = Corrected Quasi-likelihood Information Criteria, FFT = Fast Fourier Transform, Perform. Index = Performance index, RT = Reaction Time. Lower QICC values are better due to less information loss.

Table B5

GEE QIC values for Independent, AR1, Exchangeable and Unstructured correlation structures for all candidate ERD predictors and each outcome variable.

Correlation structure	Outcome variable: Performance index	Reaction time	Subjective state
Independent	43.834	91.901	71.508
AR1	41.870	44.757	40.500
Exchangeable	43.300	47.141	55.307
Unstructured	39.195	39.990	46.727

Notes: GEE = Generalized Estimating Equations, QIC = Quasi-likelihood Information Criteria, AR1 = Auto-regressive, ERD = Event-related desynchronization. Lower QIC values are better due to less information loss.

Table B6

GEE QICC values of all one-predictor ERD regression models for all outcome variables.

ERD regression model	Outcome variable		
	Perform. index	Reaction time	Subjective state
Model based on Delta band electrode Fz amplitude	6.542	9.091	25.360
Model based on Delta band electrode Cz amplitude	6.531	8.841	25.097
Model based on Delta band electrode Fz latency	6.554	8.827	25.587
Model based on Delta band electrode Cz latency	6.547	8.832	25.464
Model based on Theta band electrode Fz amplitude	6.521	8.755	25.441
Model based on Theta band electrode Cz amplitude	6.499	8.775	25.397
Model based on Theta band electrode Fz latency	6.546	8.718	25.514
Model based on Theta band electrode Cz latency	6.520	8.736	25.544
Model based on Alpha low band electrode Pz amplitude	6.458	8.662	25.723
Model based on Alpha low band electrode Oz amplitude	6.546	8.752	25.632
Model based on Alpha low band electrode PO3 amplitude	6.461	8.739	25.446
Model based on Alpha low band electrode PO4 amplitude	6.502	8.853	25.743
Model based on Alpha low band electrode Pz latency	6.537	8.774	25.418
Model based on Alpha low band electrode Oz latency	6.548	8.783	25.358
Model based on Alpha low band electrode PO3 latency	6.524	8.753	25.548
Model based on Alpha low band electrode PO4 latency	6.508	8.725	25.438
Model based on Alpha high band electrode Pz amplitude	6.466	8.805	25.802
Model based on Alpha high band electrode Oz amplitude	6.547	8.758	25.472
Model based on Alpha high band electrode PO3 amplitude	6.476	8.796	25.454
Model based on Alpha high band electrode PO4 amplitude	6.541	8.906	25.328
Model based on Alpha high band electrode Pz latency	6.542	8.755	25.544
Model based on Alpha high band electrode Oz latency	6.556	8.764	25.515
Model based on Alpha high band electrode PO3 latency	6.536	8.758	25.708
Model based on Alpha high band electrode PO4 latency	6.550	8.749	25.422

(Continued)

Table B6 (Continued)

Notes: GEE = Generalized Estimating Equations, QICC = Corrected Quasi-likelihood Information Criteria, ERD = Event-related desynchronization, Perform. Index = Performance Index. Lower QICC values are better due to less information loss.

Appendix C

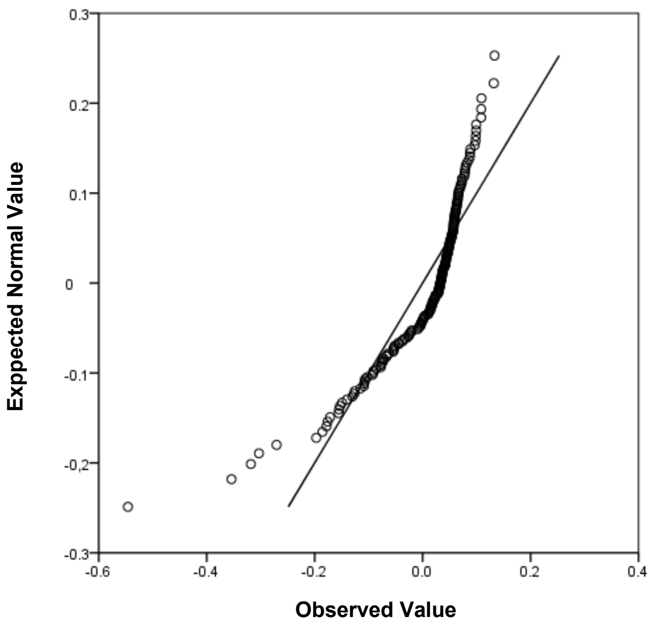

Figure C1 Quantile-Quantile (QQ)-plot of the raw residuals for Fast Fourier Transform (FFT) predictor Alpha low Pz on Performance index.

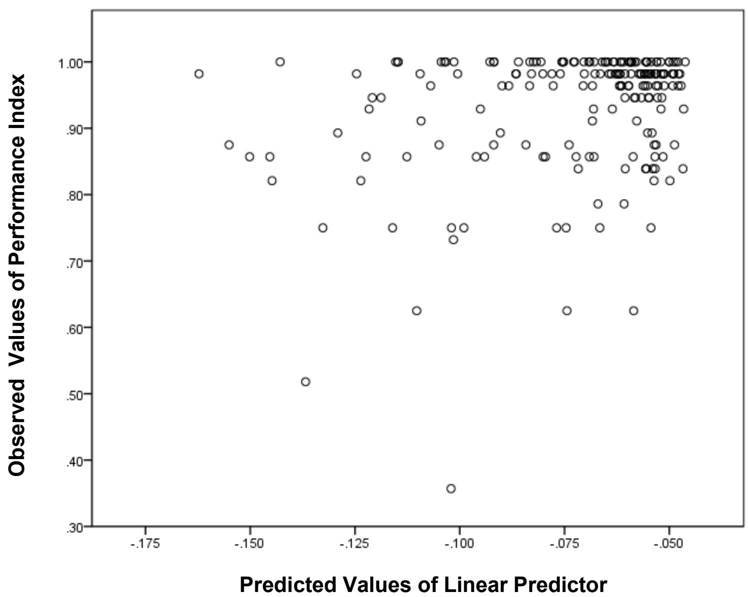

Figure C2 Scatterplot of the observed by model predicted values of Fast Fourier Transform (FFT) Alpha low Pz on Performance index.

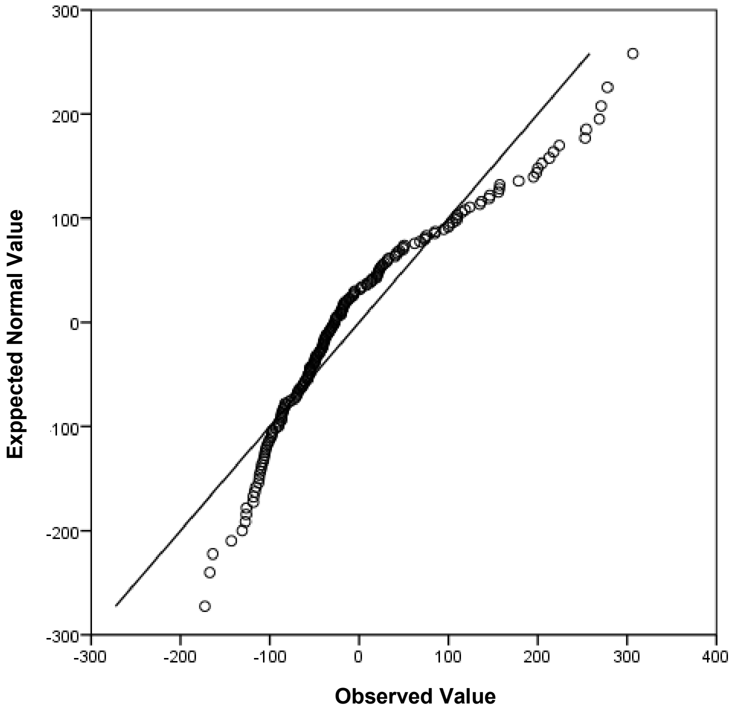

Figure C3 Quantile-Quantile (QQ)-plot of the raw residuals for Fast Fourier Transform (FFT) predictor Alpha high Pz on Reaction time.

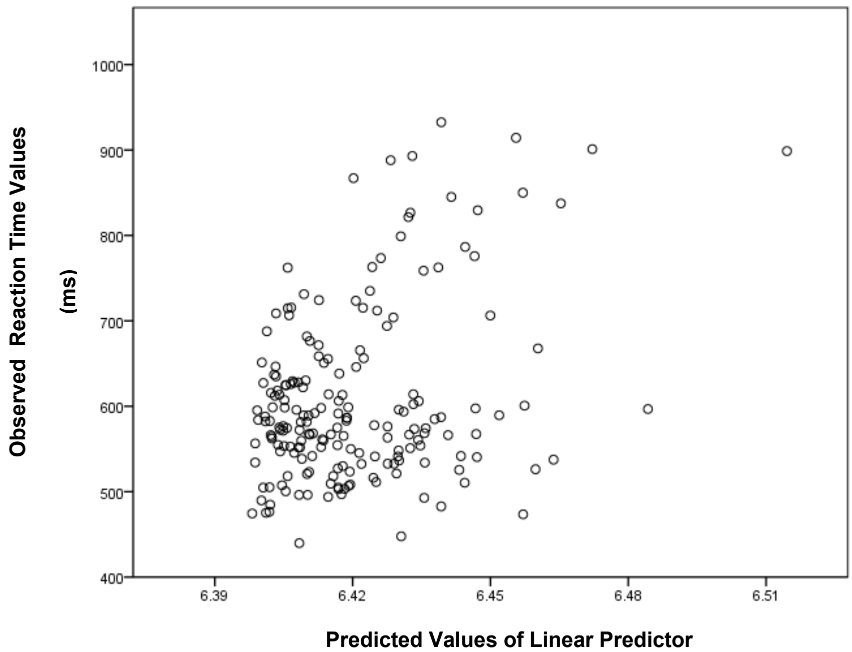

Figure C4 Scatterplot of the observed by model predicted values of Fast Fourier Transform (FFT) Alpha high Pz on Reaction time.

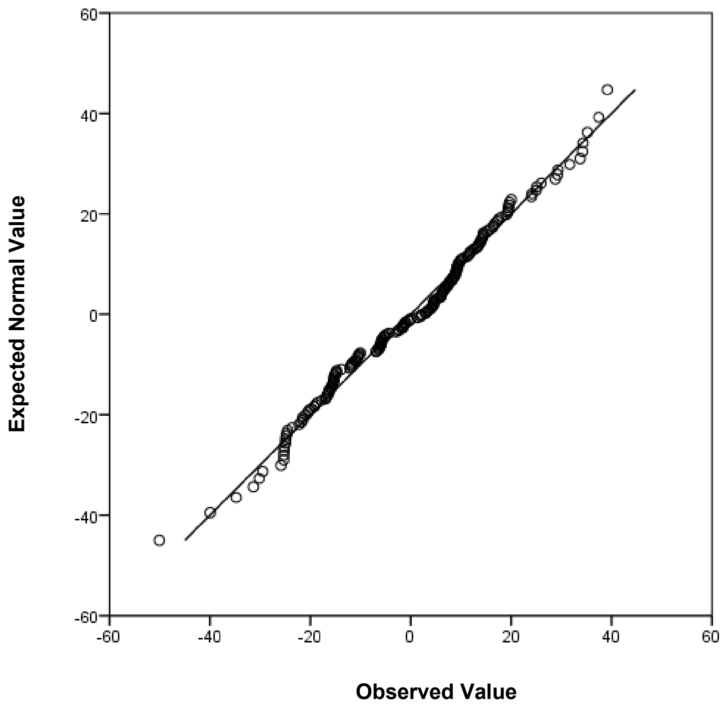

Figure C5 Quantile-Quantile (QQ)-plot of the raw residuals for Event-related potential (ERP) predictor Oz on Subjective state.

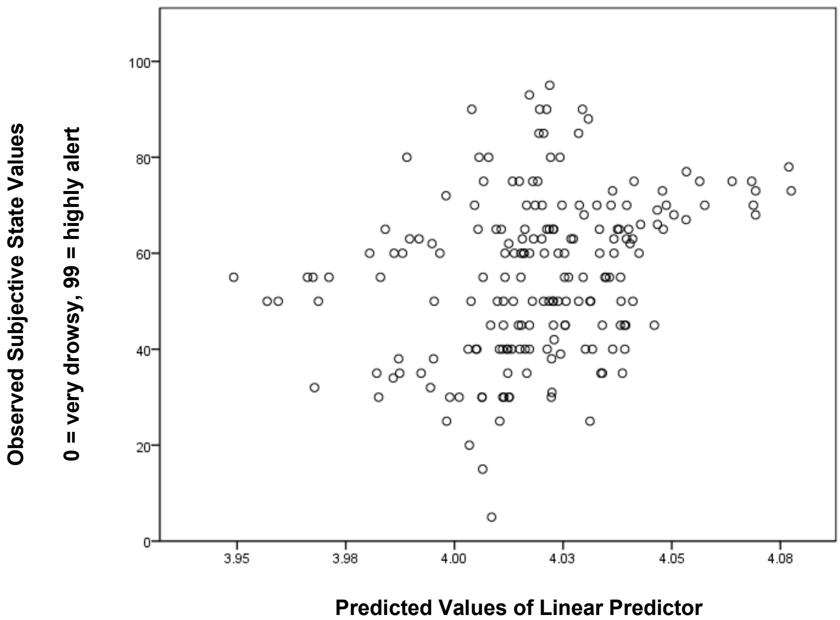

Figure C6 Scatterplot of the observed by predicted values of Event-related potential (ERP) Oz on Subjective state.