University of Tartu

Faculty of Social Sciences and Education

Institude of Psychology

Anni- Bessie Kitt

DETECTING GUILTY KNOWLEDGE WITH EEG- BASED METHODS USING MOCK CRIME SCENARIO

Master`s thesis

Supervisors: Talis Bachmann, Ph.D., University of Tartu

Inga Karton, M.Sc., University of Tartu

Running head: Detecting guilty knowledge with EEG-based methods

Tartu 2014

ABSTRACT

The aim of this study is to investigate detection of concealed information by using EEG combined with concealed information test (CIT) to reveal event-related changes in EEG. We investigate the event-related potentials (ERP) P300 and N200, which are found to be most associated with deception detection. First and foremost P300, but recent studies have shown some evidence that also N200 could be related with deception awareness.

We investigated these components using more realistic mock crime scenario, as some recent studies have implied that more realistic mock crime scenario can enhance the effects of critical stimuli. Real items in the thieving scenario and pictures of these items as stimuli were used. Furthermore, we enhance deception awareness by emphasizing in the instruction that subjects try not to get caught and give the subject constant feedback about his/her performance during the experiment. As crime scenes usually include other items in addition to the critical one (for example stolen item, murder weapon), we specified certain items as familiar stimuli in our protocol (the items from the scene which were seen, but not stolen) to investigate if the response to only seen, but not stolen items are different from the response to the stolen item.

Based on our findings we can say that more realistic mock crime scenario is related to a relatively more enhanced P300 component. Our results also support the theory that subject`s responses depend mostly on the meaning and salience of the stimuli, which are defined by the actions of the subject. The results clearly showed more pronounced P300 response for stolen item than for irrelevant or familiar items. Although it has been found that frontally measured N200 component is more pronounced in response to a critical item if deception awareness is enhanced, our results on the N200 component showed no difference in responses to stimuli with different significance.

Our findings also show that despite the fact that the P300 component effect is more enhanced in case of a more realistic scenario, the CIT protocol combined with EEG method is still not reliable enough in order to make predictions universally with any optional subject. Our results showed that detection of concealed information can be reliably detected in 29% of the subjects.

Keywords: ERP, guilty knowledge, deception detection, CIT.

KOKKUVÕTE

Süülise teadmise avastamine EEG- põhiste meetoditega kasutades kuriteo stsenaariumi

Käesoleva töö eesmärgiks on uurida süülise teadmise tuvastamise võimalikkust EEG- põhiste meetoditega, kombineerituna süülise teadmise testiga. Keskendume muutustele EEG-s esinevates sündmuspotentsiaalides P300 ja N200, mida on ka seni kõige rohkem süülise teadmise tuvastamiseks kasutatatud - eelkõige P300 komponent, kuid viimase aja tööd on viidanud ka N200 komponendi seosesele valetamise teadvustamisega.

Eesmärgiks on uurida eelmainitud komponente, kasutades realistlikuma kuriteo stsenaariumiga eksperimenti, sest mõne uurimuse tulemused on näidanud, et realistlikum kuriteo stsenaarium võib suurendada reaktsiooni kriitilisele stiimulile. Seetõttu kasutasime antud eksperimendis vargusstseenis reaalseid esemeid ning stiimulitena pilte nendest esemetest. Ühtlasi soovisime suurendada valetamise teadvustamist, rõhutades instruktsioonis katseisiku eesmärki varjata eseme varastamist ning andes katseisikule eksperimendi ajal pidevalt tagasisidet tema soorituse kohta. Arvestasime ka asjaolu, et kuriteopaikades on tavaliselt rohkem kui üks ese ning seetõttu kategoriseerisime ka oma protokollis tuttava stiimuli kategooria (ese, mida nähti kuriteopaigal, kuid mida ei varastatud), uurimaks vastuste erinevusi esemetele mida on sündmuskohal nähtud, kuid mitte varastatud, võrreldes reaktsiooniga varastatud eseme suhtes.

Käesoleva töö tulemusena saame öelda, et realistlikum kuriteo stsenaarium on seotud P300 komponendi suurema amplituudiga. Meie tulemused toetavad ka teooriat, et katseisiku vastused stiimulile sõltuvad peamiselt stiimuli tähendusest ja esilekerkivusest, mis defineeritakse läbi katseisiku tegude. Tulemused näitavad selgelt suurenenud P300 komponenti varastatud eseme suhtes võrreldes tuttava esemega. Mõned tööd on leidnud, et EEG-ga frontaalses piirkonnas mõõdetud N200 komponent on suurem kriitilise stiimuli puhul kui valetamise teadvustamine on suurem, kuid meie tulemuste põhjal ei olnud N200 komponendil erinevusi erineva olulisusega stiimulite puhul.

Kuigi P300 komponendi efekt on suurem realistlikuma kuriteo stsenaariumi puhul, ei ole süülise teadmise test kombineerituna EEG meetodiga siiski piisavalt usaldusväärne, et teha ennustusi kelle tahes üksikisiku kohta. Käesoleva töö tulemuste põhjal võib usaldusväärselt õigesti hinnata valetamist 29% katseisikutest.

Märksõnad: ERP, süüline teadmine, vale tuvastamine, CIT.

1. INTRODUCTION

Lying is a part of our society, it is a social behavior which is described as a situation when a person is trying to make believe others in something that he/she knows is not true (Ganis & Keenan, 2009). In many cases it is not vital to reveal lies or concealed information, but in some cases it might be very important to be able to detect deception. For example in case of burglary or murder investigation, it could be very useful if a reliable method for deception detection would be developed. Unfortunately, at the moment this is not the case, but there have been improvements in developing these kinds of methods and the aim of this research is also to contribute to deception detection reserach.

Researchers have been combining and developing different methods to find out what would work reliably and could be used in real life situations. Most common psychophysiological measures used with polygraph have been respiration, GSR, cardiovascular measures and electrodermal response (EDR) (e.g. Ben- Shakhar and Elaad, 2003; Carmel, Dayan, Naveh et al., 2003; Lykken, 1959). Recently reseachers have been conducting studies that are based on EEG and brain imaging. In EEG based research, the aim is to record brain potentials to reveal event-related changes, called event-related potentials (ERP) in association with deceptive behavior (Ambach, Bursch, Stark et al., 2010; Farwell, 2012; Farwell & Donchin, 1991; Luck, 2005; Rosenfeld, Hu & Pederson, 2012).

In the current study we are using a version of the Guilty Knowledge Test (GKT), recently known also as the Concealed Information Test (CIT; Lykken, 1959, 1979; Rosenfeld, 2011). CIT is a modern method of polygraph interrogation that advances psychophysiological detection of prior knowledge of crime details that would be known only by the perpetrator. It is developed to detect whether or not a person (e.g. a suspect) recognizes the importance of information, which could be known only by the guilty perpetrator. In the CIT, psychophysiological responses to critical (i.e., the so-called relevant *probe*, potentially accusation allowing or incriminating) items are compared to the responses to neutral (i.e., the *irrelevant*, contextually not significant) items whilst subjects are trying to hide or deny that they have specific contextuated knowledge of the critical items. It is posited that critical items lead to enhanced responses compared to the responses to neutral items, and in that case, concealed information knowledge is said to be detected (Ambach et al., 2010; Farwell, 2012; Rosenfeld, 2011).

In addition to the two item protocol, some studies have used a three item protocol, which includes an additional stimulus - a target (Farwell & Donchin, 1991; Mertens & Allen, 2008; Rosenfeld et al., 2012; Rosenfeld, Soskins, Bosh et al., 2004). The target is basically an irrelevant item from the point of view of desired concealment, but the subject is assigned to make a unique button response which would be different from the responses to irrelevant and probe items. The target is meant to be used to keep the subject`s attention to unpredictable stimulus that is randomly presented. However, there is no clear consistency in the ERPs associated with targets as the responses to targets may be sometimes significantly less distinctive than responses to irrelevants (Gamer & Berti, 2010; Seymour, Seifert, Shafto et al., 2000). In some studies the probes and targets have produced similar ERPs (Farwell & Smith, 2001) but in some cases targets have produced larger P300 amplitudes than probes (Mertens & Allen, 2008). Therefore, in the present work we do not use the protocol that includes targets as a special category of stimuli.

The approach used in this study is more similar to the classical two-item CIT protocol, but involves two types of irrelevant stimuli. In this study we use a protocol that includes (i) *probe*, (ii) *irrelevant* and (iii) a different kind of irrelevant, which was termed *familiar*. We labelled the stimulus familiar as it represents the stimuli which were present during the enactment of the crime. As our aim is to make the mock crime situation more realistic and usually there are more than one item on the scene, we differentiated a familiar stimulus (seen during the enacment of crime) from neutral (no prior exposure) stimuli in the protocol. The category of familiar stimuli allows to have one useful comparison condition against the probe (critical) item condition. By this we can asses the relative importance of the factor of simple visibility and/or visual memory with regard to the significance of the seen stimuli.

In deception detection studies the ERP component P300 is used most often, because P300 is typically elicited by the presentation of meaningful, rare stimuli, randomly presented among the non-meaningful stimuli. The P300 indicates a special status of an item in cognitive processing of a recognized or infrequent stimulus and is therefore suitable for detecting concealed information (Elaad, 2000). P300 is a positive-going wave which is largest parietally, medium centrally and smallest frontally. The peak of this component has usually latency of 300 - 800 ms measured from stimulus onset (Abootalebi, Moradi & Khalilzadeh, 2009; Farwell & Donchin, 1991; Luck, 2005; Wu, Hu & Fu, 2009). The P300-based CIT utilizes P300 relative amplitude as an index of actual recognition of concealed information (Abootalebi et al., 2009) because it is known that if a concealed information related critical stimulus is presented, the P300 in response to this stimulus is enhanced compared to neutral stimuli (Ambach et al., 2010; Farwell, 2012; Rosenfeld, 2011; Rosenfeld et al., 2012).

Another ERP component which we are investigating in this study is N200 component. Recent studies suggest that concealed information in CIT is related to this component by virtue of critical stimulus eliciting larger frontal N200 than the neutral stimuli (Gamer & Berti, 2010; Hu, Wu & Fu, 2011; Wu et al., 2009). Hu, Pornpattananangkul and Rosenfeld (2013) found that probes elicited a larger N200 compared to irrelevant stimuli only when guilty participants were highly aware of their deception but not otherwise. They also suggested that in case of high deception awareness subjects monitored their responses more carefully compared to low awareness situation. This is supported by a study where researchers found that frontally increased N200 is also found in cognitive control tasks involving conflict monitoring (Folstein & van Petten, 2008). It has also been hypothesized that as feedback hightened deception awareness among guilty participants, they might have been more engaged in monitoring their responses and performance (Folstein & van Petten, 2008; Wu et al., 2009).

An important aspect regarding the use of EEG is that it can be time-consuming because a lot of single trials are necessary to get a reliable result with EEG; conducting more trials allows to get more clear and systematic ERP-s (Luck, 2005). Therefore it is imporant to make as much single trials as possible, but on the other hand, it is exhausting for the subject, so eventually the tiredness, lack of motivation and attention can affect the results (Woodman, 2010). Using less trials per subject and a larger number of subjects cannot be used in this paradigm, because single subject level assessment is very important in this case, therefore efforts should be made to try to get as many trials as possible from a single subject. The experiment should be designed in a way that it is possible to conduct as many trials as needed without compromising subject`s attention and motivation and also granting that the experiment cannot be too long (for example by dividing trials into smaller blocks of trials) (Luck, 2005; Woodman, 2010). It has been suggested that at least around 30-33 trials (eye-blinks and other artifacts removed) in each condition should be available to make any further analysis (Dietrich, Hu, Rosenfeld, 2014; Rosenfeld, Hu, Labkovsky et al., 2013).

There are several aspects that can moderate the efficiency and reliability of a CIT. As the aim of this research was to make the experiment more realistic, we also considered some aspects to enhance reality-likeness of the situation. We made the mock crime scenario more realistic by using real objects for the thieving episode in the mock crime scenario and used pictures of these objects as stimuli. Mock-crime scenario has been used in a couple of studies: some of them have used a text-based interrogation in a CIT (Jokinen, Santtila, Ravaja et al., 2006); written words have been used as stimuli (Hu et al., 2013, Winograd & Rosenfeld, 2011); and some of the researchers have used pictures as stimuli (Matsuda, Nittono, Allen, 2013; Sip, Carmel, Marchant et al., 2013). Cutmore, Djakovic, Kebbell and Shum (2009) assessed the difference between picture and word stimuli in their study and found that using pictures of objects may be more effective than using words.

Another way to make the experiment more realistic is to increase deception awareness (i.e., experienced significance). This can be done with instructions that emphasizes the importance of not getting caught and motivating subjects to hide their crime-related knowledge throughout the experiment. Also, letting a subject to keep/hide the stolen item with him/her is useful in this respect. One possibility is to give feedback, especially continous feedback during the experiment, to enhance deception awareness and motivation to hide knowledge of the critical item (Nahari & Ben- Shakhar, 2011; Rosenfeld, Hu & Pederson, 2012; Winograd & Rosenfeld, 2011).

Some recent studies have found that salience of the stimuli is important in deception detection as it moderates the subject`s sensitivity to the test (Carmel et al., 2003; Jokinen, et al., 2006; Nahari & Ben- Shakhar, 2011; Verschuere, Rosenfeld, Winograd et al., 2009). Jokinen and colleagues (2006) hold that salience increases through subject`s actions or remembering details through their own actions. Therefore he suggests that it is important to concentrate on the actions of the perpretator and details related to perpetrator`s actions, which helps create more effective questions. The importance of questions was also raised by Ambach, Dummel, Lüer, and Vaitl (2011), who compared the question style („did you see?“ vs „did you steal?“) with two groups of critical stimuli (only seen vs seen and stolen items). They found that both types of critical items, regardless of the question type, elicited different physiological responses (skin conductance, electrocardiogram, respiration, and finger pulse) compared to neutral items. Although only for the stealing question, responses to stolen objects were different from responses to objects that were only seen, but not stolen.

It is widely known that subjects might use countermeasures to defeat the polygraph test, so researchers have thought of the ways to make it more difficult for the subject to use countermeasures. In CIT-based studies one way to do this is to make the task cognitively more difficult; for example, an efficient possibility is to increase the number of neutral stimuli (Hu, Hegeman, Landry et al., 2012; Rosenfeld et al., 2013).

In this study our aim is to assess the reliability of the P300 component on a single subject level as many previous researchers have done (Abootalebi et al., 2009; Allen et al., 1992; Farwell & Donchin, 1991; Mertens & Allen, 2008; Rosenfeld, Biroschak & Furedy, 2006; Rosenfeld, et al., 2004). The single-subject level assessment is especially important for practical reasons allowing us to understand how effective the method is on an individual level because in forensic practice individual assessment is the only possible way. Different methods, such as bootstrap analysis (Rosenfeld, 2011; Rosenfeld et al., 2006) or signal detection theory (ROC; Hu et al., 2013), have used individual P300 responses to detect deception. In the current study we use a bootstrapped amplitude difference (BAD) method, which we do hope is not so bad at all.

1.1. Research question and hypothesis

Previous research has shown that the first and foremost component related to detecting concealed information is P300 (Ambach et al., 2010; Farwell, 2012; Rosenfeld, 2011). Although recent studies have also shown that variability in an additional component, N200 could be related to concealed information -- mostly through deception awareness, response conflict and monitoring responses (Gamer & Berti, 2010; Hu, et al., 2011; Wu et al., 2009). In this study we are investigating these components using a more realistic mock crime scenario by using real items in the mock crime scene and pictures of them as stimuli. Furthermore, we enhance motivation to deception and its awareness in the experimental instruction by emphasizing the need not to get caught, by letting a subject to keep the stolen item nearby and hide it and by informing the subject that he/she will be given constant feedback about his/her performance during the experiment, which should increase N200 (Wu et al., 2009). As crime scenes usually include other items in addition to the critical one (i.e., the stolen item), we differentiated the category of familiar stimuli in our protocol and examine responses to familiar stimuli also; this is in order to see if there are any differences compared to responses to critical or neutral stimuli.

Based on previous research we hypothesize that a critical stimulus would elicit increased P300 compared to the neutral and familiar stimuli. We also hypothesize that a critical stimulus will similarly elicit increased N200. Additionally, we will investigate the reliability of the P300 component on a single subject level.

2. METHOD

2.1. Subjects

Twenty nine volunteers participated in the experiment, all subjects were healthy and had normal or corrected to normal vision. All participants gave written informed consent before participating in the study. The experiments were approved by the Research Ethics Committee of the University of Tartu and were conducted according to the principles set in the Declaration of Helsinki. All subjects received monetary compensation for participation, in the amount of 5 euros. The data of five male and seven female participants was excluded due to measurement problems and extensive blink artifacts. Seventeen healthy volunteers (12 females and 5 males, mean age 25.52, SD = 3.16; range 23-36) remained for further analysis.

2.2. Behavioral task and procedure

The task was similar to the various tasks used in the format of GKT and CIT in that critical (probe), neutral (i.e. unfamiliar from the experimental behavioral context) and familiar (from the experimental behavioral context) stimuli were used and responses to these three stimuli classes were compared. There were 1 critical, 3 familiar and 4 neutral stimuli. Experiment started with a thieving episode (a mock crime scenario) with participating subjects stealing real objects. The purpose of the experiment as explained to the subject was to discover „stealing“ using EEG and computer. The subject was instructed that his/her goal was to always deny „stealing“ the item. The subject was also notified that during the experiment he/she will also receive feedback (if computer has „detected“ lying) from the program. The subject was motivated to hide the “crime” related knowledge.

In the experiment eight objects were used: calculator, car key, ring, wallet, photo camera, wristwatch, earrings, cell phone. For the stealing episode subjects entered a room where a random selection of four out of eight objects were placed on a table. The subjects were asked to „steal“ one of these objects and put it in a bag (the bag was given before the experiment by the experimentator). On half of the cases eyewitness was present (to further enhance deception awareness). The eyewitness was working with a computer behind a desk that was in the same room where the mock-crime took place, so the subjects understood that the eyewitness could see what item was taken. Note, however, that we did not find any effects related to the presence of an eyewitness and thus we disregard this factor in our subsequent analyses. After the stealing episode an EEG cap was fitted to the subject's head and the main part of the experiment begun.

Subjects were seated at a distance of 100 cm from the computer monitor (SUN CM751U 1024 x 768 pixels, 100 Hz refresh rate). During the experimental task photographs of the eight items were presented foveally in random order. The experiment was divided into 6 blocks with 100 trials in each block. Thus, each of the stimuli were presented 75 times in total. The photographs were further complemented by names of the items printed with high-contrast dark letters under the photographs. The digital photographs from all of the eight objects were taken with digital photo camera Pentax Optio M85. The photographs were taken on the same table on which the subjects were told to „steal“ the object, therefore the background for the pictures was the view of the table, patterned light wood. The words were added to the digital photographs in Picasa 3.9. The stimuli were presented on a light background with a luminance of 86 cd/m^2. Mean luminance of the stimuli was 38 cd/m^2 (SD = 13.9 cd/m^2). Each trial begun with presenting a fixation cross in the middle of the screen. After 300 ms the fixation cross was replaced by a blank screen for a random duration between 1000 – 1500 ms. The duration of the stimulus that was presented thereafter was set at 1400 ms. After each stimulus a question appeared on the screen: „Did you take this object? YES/NO“; the sentence thus included also answer options. The words were written in black letters on a white background. The question and response options remained on the screen until the subject gave a response (it was possible to respond as soon as the sentence appeared on the screen). Responses were typed in by the dedicated response keys on the computer keyboard, held on the lap by the subject. Subject was instructed to give an answer by pushing the left arrow on the keyboard for „YES“ and right arrow on the keyboard for „NO“; subject was also instructed to give the answer verbally (to enhance deception awareness). After responding, subjects initiated the next trial by pressing the spacebar on the keyboard. Timing of the trials was free so that subjects were able to blink eyes if necessary (which reduces eye-movement artefacts in EEG recording). Feedback appeared on the screen after every 43 and 79 trials. As the feedback did not occur very often, the feedback remained the same each time it occurred: „The program has detected lying, try to be more careful and hide more efficiently!“. The duration of the feedback screen was set at 5000 ms. The purpose of the feedback was to keep the subject’s attention, motivation and enhance deception awareness.

2.3. EEG and data analysis

We used the Nexstim eXimia EEG-system with a 60 carbon electrodes cap (Nexstim Ltd., Helsinki, Finland) for EEG recordings. In both experiments, the impedance at all electrodes was kept below 10 kΩ. The EEG signals were referenced to an additional reference electrode placed on the forehead and sampled at 1450 Hz. All signals were amplified with a gain of 2000 and filtered with a hardware based bandpass filter of 0.1–350 Hz. The vertical electrooculogram (VEOG) was recorded via two additional electrodes placed above and below the participants' left eye. All recorded EEG data were analyzed with Fieldtrip (http://fieldtrip.fcdonders.nl; version 14-12-2013), an open-source MATLAB toolbox.

EEG was recorded from 15 electrodes: frontal (F5, F6, Fz), frontocentral (FC3, FC4, FCz), central (C3, C1, C2, C4, Cz), parietal (PO3, PO4, Pz). After the initial recording, data were high-pass and low-pass filtered (with zero phase shift Butterworth filters of 0.1 and 30 Hz, respectively) and segmented into trials from -200 ms to 1000 ms relative to stimulus onset. The data were manually checked for any artifacts, including eye movements and blinks. All trials contaminated by artifacts were discarded from further analysis. Data were baseline-corrected with a 100 ms window prior to the stimulus onset and linear trends were removed from the data.

Average ERP potentials were computed for each subject in each condition and for each electrode. The activity from single electrodes was pooled into two regions of interest (ROI's). The ROI's were frontal (electrodes FC3, FCz, FC4), and parietal (electrodes PO3, Pz, PO4). The experimental conditions were neutral (stimuli that were not encountered previously during the mock crime), familiar (stimuli that were encountered during the mock crime) and critical (the stolen stimulus typically labelled as probe in CIT). On average, the following number of trials were available: for the neutral condition: mean = 154.35, SD = 13.59; for the familiar condition: mean = 114.76, SD = 8.87; for the critical condition: mean = 40.94 , SD = 7.40. Mean amplitude was used for the analysis of ERP components. First, the relevant component peaks were identified from a grand average ERP over all conditions, all ROI's and all subjects. Second, the duration of each component was determined by visual inspection. The length of the period for mean amplitude was then set at half the duration of the respective component and centered at it's peak. Specifically, the following criteria were applied in the experiment: for N2, 271 ± 20 ms; for P3, 451 ± 75 ms. Mean amplitude was used instead of other possible measures (such as peak amplitude, for example) because it is not biased by unequal signal-to-noise ratios across conditions (Luck, 2005) and thus the different number of available trials across our experimental conditions is not a serious problem.

2.4. Statistical analysis

Statistical analyses were performed with R (version 3.0.3), a freely available and powerful statistical programming language. Repeated measures analysis of variance (ANOVA) was used to assess the effects of our experimental conditions. If the sphericity assumption was violated according to Mauchly's test for sphericity, p-values were corrected with the Greenhouse-Geisser method. Only the corrected p-values are reported. As recommended by Bakeman (2005), generalized eta-squared measure is used to report effect sizes of our ANOVA results. Planned comparisons and post-hoc contrasts were carried out via dependent samples t-test. Post-hoc contrasts were corrected with the Holm–Bonferroni method. Unless indicated otherwise, only the corrected p-values are reported. Cohen's d is reported as an estimate of effect size for the dependent samples t-tests.

To assess the reliability of this experimental effect within each participant, a bootstrapping test was performed as recommended by Rosenfeld et al. (2006). A 100 iterations of the standard ERP extraction procedure were performed for each subject. On each iteration, 25 trials were randomly chosen from all available trials in the critical and in the neutral condition. These trials were averaged according to their condition and the mean P300 amplitude was calculated using the above described criteria. Finally, the difference in mean P300 amplitude for the critical and the neutral condition on a given iteration was computed. Thus, after a hundred iterations we obtain for each subject a percentage of iterations where the mean P300 amplitude of the critical condition is higher than the mean P300 amplitude of the neutral condition. Or put differently, we obtain a percentage of iterations where the previously found experimental effect is evident for a given subject.

3. RESULTS

The aim of this experiment was to study how the significance of crime related items influences the amplitude of the N200 and the P300 components in a mock crime scenario, where real objects were stolen during the mock crime and pictures of those objects were used as stimuli in a situation where efforts were made to enhance deception awareness and motivation. Most importantly, we distinguished between three types of stimuli with a varying degree of their significance in the experiment. There were critical stimulus (the item actually stolen by the subject), familiar stimuli (items that were present during the crime episode, but not stolen by the subject) and neutral stimuli (subjects had no prior knowledge associated with these items in the context of mock crime). We expected to find ERP differences between the critical compared to the neutral and familiar stimuli conditions. To assess the neural correlates of concealed information in this experiment, we compared mean N200 and P300 amplitudes across our experimental conditions. We also assessed the reliability of the P300 signature on the single subject level.

First, a two-way repeated-measures ANOVA with the factors electrode group (frontal and parietal) and stimulus type (critical, familiar, neutral) was performed to assess differences in mean P300 amplitude. The condition-specific ERP waveforms can be seen in figure 1. The main effect of electrode group was significant ($F(1,16) = 71.7$, $p = 0.00000026$; $\eta G^2 = 0.34$). This was due to the much lower mean amplitude of frontal electrodes ($m = 2.59$, $SD = 2.82$) compared to parietal electrodes ($m = 7.07$, $SD = 4.48$) (Table 1). The main effect of stimulus type was also significant ($F(2,32) = 11.2$, $p = 0.003$; $\eta G^2 = 0.08$). There was no interaction between electrode group and stimulus type ($F(2, 32) = 1.7$, $p = 0.21$; $\eta G^2 = 0.004$).

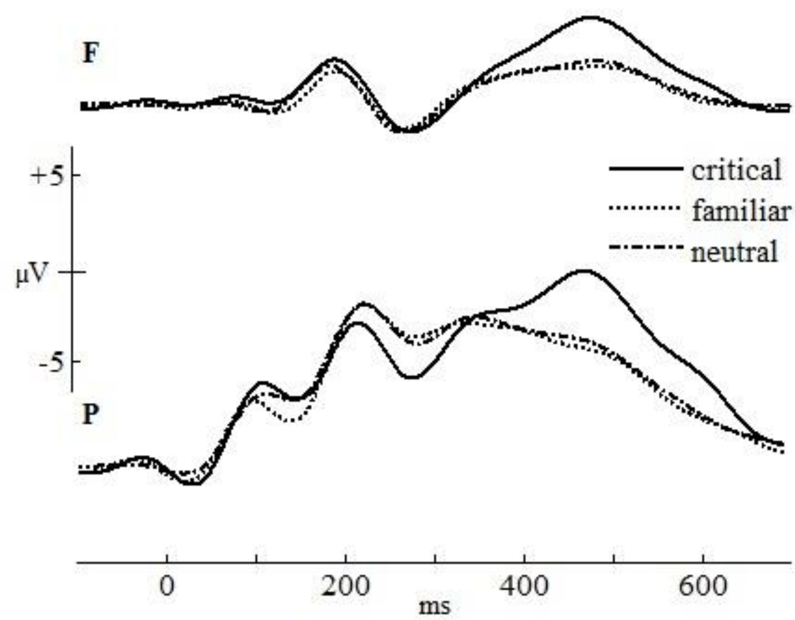

Figure 1. ERP's obtained in the critical, familiar and neutral conditions, recorded from frontal (F) and parietal electrodes (P). (For visually assessing the absolute potential levels for F and P conditions the common voltage scale should be extrapolated so as to align the 0 microvolts level with pre-stimulus baseline.)

Planned comparisons were carried out to investigate which conditions differed significantly from each other in terms of the mean P300 amplitude which they elicited. For the parietal electrode group, there was a significant difference between the critical and the neutral condition ($t(16) = 3.45$, $p = 0.003$, $d = 0.84$) and the critical and the familiar condition ($t(16) = 3.4$, $p = 0.004$, $d = 0.83$). The difference between the neutral and the familiar condition was not significant ($t(16) = -1.03$, $p = 0.32$, $d = 0.25$). For the frontal electrode group, the differences were similar: there was a significant difference between the mean P300 amplitude of the critical and the neutral condition ($t(16) = 2.62$, $p = 0.019$, $d = 0.64$) and the critical and familiar condition ($t(16) = 2.57$, $p = 0.02$, $d = 0.62$). The difference between the neutral and the familiar condition was not significant ($t(16) = -0.4$, $p = 0.69$, $d = 0.1$). Figure 2 and table 1 show the mean P300 amplitudes and significant differences between the experimental conditions.

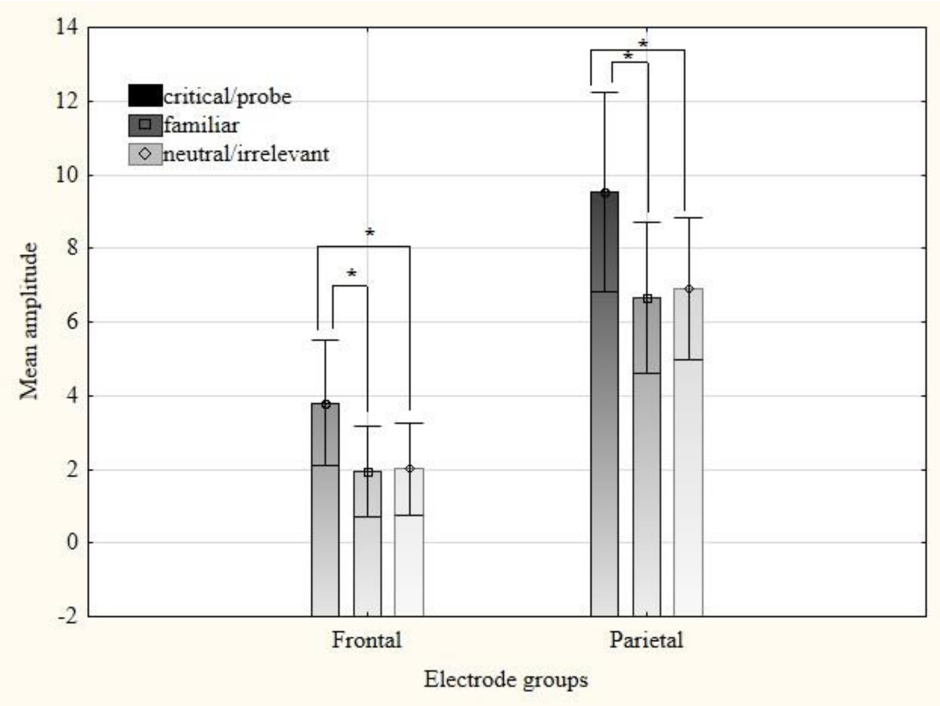

Figure 2. Paired t-test differences for the comparison of the effects of critical, familiar and neutral stimuli in frontal and parietal mean area ERP recordings. * $p<0.05$.

Second, a two-way repeated-measures ANOVA with the factors electrode group (frontal and parietal) and stimulus type (critical, familiar, neutral) was performed to assess any potential differences in mean N200 amplitude. The ERP waveforms can be seen in figure 1. The main effect of electrode group was highly significant ($F(1,16) = 122.5$, $p = 0.0000000066$; $\eta G^2 = 0.55$). There was no main effect of stimulus type ($F(2,32) = 2.6$, $p = 0.12$; $\eta G^2 = 0.02$), but the interaction between electrode group and stimulus type was significant ($F(2,32) = 11.7$, $p = 0.001$, $\eta G^2 = 0.02$).

Table 1. Paired t-test means and standard deviations for ERP amplitude differences between stimuli conditions.

Area	Electrodes	Critical/Probe		Familiar		Neutral/Irrelevant	
		mean	SD	mean	SD	mean	SD
N200	Frontal (F)	-1.31	3.27	-1.15	2.60	-1.30	2.76
	Parietal (P)	4.84	4.98	6.98	3.58	6.74	3.28
P300	Frontal (F)	3.80	3.32	1.94[b]*	2.38	2.02 [a]*	2.43
	Parietal (P)	9.53	5.28	6.67[b]**	3.98	6.91[a]**	3,72

Notes: paired t-test p-values: *$p<0.05$, **$p<0.01$; Holm–Bonferroni corrected p-value: °$p<0.05$, °°$p<0.01$
[a] Neutral<Critical
[b] Familiar<Critical

Post-hoc paired t-tests were carried out to find the sources of the significant interaction effect. After correcting for multiple comparisons, however, no significant differences between our three stimulus conditions can be reported. For the frontal electrode group, the critical and neutral ($t(16) = -0.02$, $p = 1.0$; $d = 0.005$), the critical and familiar ($t(16) = -0.28$, $p = 1.0$; $d = 0.07$) and the familiar and neutral conditions ($t(16) = 0.68$, $p = 1.0$; $d = 0.16$) did not differ in terms of their mean N200 amplitude. The same was true for the parietal electrode group. There were no differences between mean N200 amplitudes of the critical and the neutral ($t(16) = -2.54$, $p = 0.11$; $d = 0.62$), the critical and familiar ($t(16) = -2.67$, $p = 0.1$; $d = 0.65$) nor the familiar and neutral conditions ($t(16) = 0.86$, $p = 1.0$; $d = 0.21$). Table 1 shows the mean N200 amplitudes and paired t-test results between our experimental conditions.

From the results we can see that only the P300 component exhibits clear systematic amplitude differences in response to critical items, it also differentiates between the critical items on the one hand and neutral and the familiar items on the other hand. This effect is equally clear on the parietal and the frontal electrode group, although the effect sizes are large for the parietal and medium for the frontal electrode group. Even though we observed a significant electrode group by stimulus type interaction for the mean N200 amplitude, the effect size of this result was rather small and post-hoc tests could not confirm any specific differences between the three types of stimuli.

To determine whether the critical vs neutral difference in mean P300 amplitude is consistent within single subjects the bootstrap method (Rosenfeld et al. 2006; see Methods for more details) was used. Table 2 shows the results for all 17 subjects. For each subject the percentage of iterations where the mean P300 amplitude of the critical condition exceeded the mean P300 amplitude of the neutral condition is shown. Only five (29%) participants reached the necessary 90% criterion for reliable deception detection in Experiment 2. Figure 3 gives some examples of subjects with the highest and with the lowest hit rates on the bootstrap test.

Table 2. Bootstrap analysis results (Experiment 2)*

Subject	Parietal	Frontal+Parietal
1	30	39
2	1	1
3	66	65
4	61	76
5	90	97
6	48	17
7	19	76
8	14	5
9	96	95
10	0	0
11	95	92
12	31	22
13	60	71
14	100	100
15	10	9
16	98	99
17	0	0
Corr class (90) %	29	29
Mean %	48	51

*Note: Percent of iterations for which critical was higher than neutral in mean amplitude. In the first case only mean P300 amplitude of the parietal electrode group was used for the bootstrap test. In the second case the mean over both electrode groups was used.

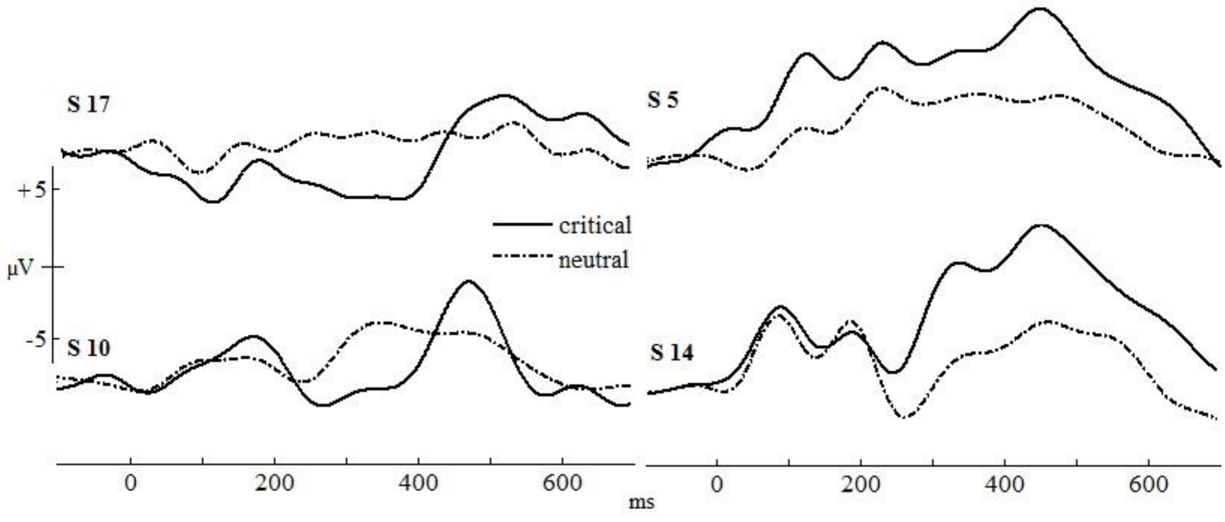

Figure 3. Examples of averaged (10 Hz) ERP's of individual subjects with high critical > neutral index (S5, S14) and low critical > neutral index (S17, S10).

4. DISCUSSION

The aim of this study was to reserach the response to different crime related items in a mock crime situation, where deception awareness and motivation was enhanced, the subject was given constant feeback and the situation along with the stimuli were more realistic. Responses to the items with different crime-contextual significance level were compared by the mean amplitudes of N200 and P300 ERP components. We expected to find significantly increased effect in case of critical condition compared to neutral condition. We also expected to find significantly increased effect in response to a critical condition compared to the familiar condition (which represented items seen from the same mock crime scene, but which were not stolen). In addition, we assessed the reliability of the P300 component on a single subject level.

Our first hypothesis was confirmed by the results. The P300 response to a critical stimulus is clearly enhanced compared to the neutral condition. This effect is equally clear on the parietal and the frontal electrode group although the P300 effect size is large for the parietal and medium for the frontal electrode group. Our results support the findings that have been reported by several previous studies (Abootalebi et al., 2009; Ambach et al., 2010; Farwell & Donchin, 1991; Wu et al., 2009). This also suggests that the responses towards a critical stimulus are clearly stronger than responses to neutral stimuli in a more realistic mock crime scenario. Furthermore, previous studies have shown that more realistic stimuli have shown better results in CIT studies (Ambach et al., 2010; Cutmore et al., 2009). It has also been implied that in case of deception the cognitive demand is higher, so therefore the memory work load increases and this might also increase the P300 response (Verschuere et al., 2009). As in this case the mock crime was more realistic, deception awareness was enhanced and there were seven stimuli in addition to the critical stimulus, the cognitive demand was increased, which would also confirm the clear results on P300 component.

The importance of more realistic mock crime is also supported by a recent study by Karton, Kitt, Rutiku and Bachmann (2014, submitted paper) where we compared two similar CIT studies using mock crime scenario. One experiment was the experiment described in this study and the other was a similar mock crime experiment, but less effort was made to enchance deception awareness (for example there was no feedback during the experiment), cards with words written on them were used in the mock crime and words were used as stimuli. The results in comparing these two experiments clearly showed more pronounced effects in the more realistic mock crime scenario than in the more artificial experiment where information processing was mediated by a more abstract denotation of the stimuli items. The added value of this study was that the results were obtained in the same lab environment, using the same experimental equipment and concentrating on the same ERP sigantures which allowed more reliable comparison between different types of mock crime.

We also confirmed that the P300 response to a critical stimulus is clearly enhanced compared to the familiar stimulus. The results show that the response towards familiar stimulus (it represents the stimuli which were present during the enactment of the crime, but were not stolen) significantly differs from the response to critical stimulus, but does not differ from neutral stimuli. On the basis of this knowledge we can hypothesize that CIT does not show recognition of a seen item from the crime scene, but the item has to have specific critical meaning to the subject for the enhanced response to be revealed.

The results are encouraging as they confirm that CIT is more than just a test for detecting recognition of familiar items, hence it differentiates familiar and irrelevant stimuli from the critical stimulus. Therefore it can be recommended that gathering as much information as possible during the investigation makes it possible to make the CIT maximally effective. This includes using more realistic stimuli and focusing on the quality of the questions by focusing them on the perpetrator`s actions as the critical item salience increases through perpretator`s actions (Jokinen et al., 2006). Combining this knowledge with the fact that pictures are more effective stimuli compared to abstract symbolic format of presentation, it can be a useful information for future research. While in this study we did not focus on question analysis used in the CIT, in subsequent research it would be very important to study further the significance of questions used in CIT. Ambach and colleagues (2011) also confirm the importance of the test question as they found in their study that the response to the stolen item was enhanced compared to the response to the item only seen if the question that was asked was about stealing the item („Did you steal..?“), but not when the question was about seeing the item (Did you see?“).

Bearing in mind that the aim of this study was to be more realistic in conducting the mock crime scenario we gave continous feedback and tried to enhance deception awareness with the instructions also. Verschuere et al. (2009) did not find that deception awareness and feedback would enhance P300 in case of deceptive subjects, but as they were using the subject`s name as the probe, the stimuli was very salient and rehearsed, also they did not use continous feedback in their experiment. Rosenfeld and colleagues (2012) obtained more clear-cut results in their study using information rehearsed to a lesser extent as stimuli and also using continous feedback. In the present study the deception awareness might have been more effective due to the fact that the information was not rehearsed before the experiment and we used continous feedback throughout the experiment.

Our second hypothesis was not confirmed as the results for N200 were not what we expected. We hypothesized that frontally measured N200 would be more pronounced in response to a critical stimulus compared to a neutral stimulus. Previous studies (Hu et al., 2013; Rosenfeld et al., 2013) have showed that higher deception awareness and motivation should evoke the response to a critical item which is more pronounced compared to the response to neutral stimuli. In our case, the results showed no differences in the responses to critical stimulus compared to neutral or familiar stimuli.

Wu and colleagues (2009) found enhanced N200 with deceptive response, especially in case of forced response (in our case it was also obligatory to steal one item and later deny stealing it). They imply that their results are consistent with N200 being an index of conflict detection which is related to increased workload and an enhanced control process. Hu and colleagues (2009) also contended -- based on their results -- that N200 is related to conflict detection, stimuli categorization and response selection. Hu and colleagues (2013) noted that since the N200 was enhanced only in the high awareness condition, the subjects monitored their responses to the critical stimulus to a greater extent. One possible explanation for why the effect did not occur in our study even though efforts were made to enhance deception awareness is that we had one question about stealing an item from the table, so the subject had to say „NO“ after each stimulus. If the stolen item was presented, the subject had to deny taking the object and in case of other seven items, the subject replied also „NO“ and was honest. We can hypothesize that since the subject had to say „NO“ in response to each stimulus and always pushed the same button, they did not have to use enhanced cognitive control and monitor their response very thoroughly, because the reply was always the same and may have been automatized. This discussion also refers to the aforementioned aspect about the importance of the question and as it seems, response versions also. We suggest that different aspects which could moderate the N200 component should be further investigated, because this component has not been studied very thoroughly and different results have been found.

In practice, the only possibility to detect concealed information is to assess the results on an individual level. This is also the reason why our interest was to investigate how reliably can we use the P300 marker on a single subject level. The amount of correctly detected deceptions using different methods has been quite high in earlier studies, ranging between 85% and 95% (Rosenfeld et al., 2006; Rosenfeld et al., 1991; Farwell & Donchin, 1991; Allen et al., 1992). However, more recent studies have reported detection rates varying from 27% to 86% (Abootalebi, et al., 2009; Mertens & Allen, 2008; Rosenfeld, et al., 2004) or even have found no diagnostic power of the P300 at all (Gamer & Berti, 2010). We conducted a bootstrap analysis and showed that detection of concealed information can be highly reliably detected on 29% of the subjects in our study. This result suggests that even though the response to a critical stimulus is significantly more enhanced compared to neutral stimulus on a group level and the effect sizes might be larger than in some more artificial experiments, it does not mean that the method is absolutely reliable and universally applicable on an individual level. On the other hand, in an informal analysis we provided a group of expert evaluators the set of individual ERP recordings and asked to retrospectively predict what were the critical items. In case of this approach, the hit rate was close to 80%. So it may be that in principle the ERP based deception detection can have more potential in it, but to show this reliably requires further versatile research.

To conclude, we can say that more realistic mock crime scanrio is related to relatively more enhanced P300 component. Our results also support the findings that subjects' responses depend mostly on the meaning and realistic salience of the stimuli, which are defined by the actions of the subject: the results clearly showed more pronounced P300 response for stolen item than for merely seen or irrelevant items. Although it has been occasionally found that frontally measured N200 component is more pronounced in response to a critical item if deception awareness is enhanced, our results showed no difference in responses to stimuli with different significance. On the account that our results showed that concealed information can be reliably detected with 29% of the subjects, we can say that despite the fact that the P300 component effect is more enhanced in case of a more realistic crime scenario, the CIT combined with EEG method is still not reliable enough make predictions universally with all or majority of the subjects. On the other hand, provided that a particular subject happens to be an individual predisposed to ERP selectively sensitivity to critical items, with that particular individual the ERP based CIT test may be reliable enough.

5. REFERENCES

Abootalebi, V., Moradi, M. H., Khalilzadeh, M. A. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Computer Methods and Programs in Biomedicine*, *94*, 48–57, doi:10.1016/j.cmpb.2008.10.001.

Allen, J. J., Iacono, W. G., Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology, 29*(5), 504-522.

Ambach, W., Bursch, S., Stark, R., Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *International Journal of Psychophysiology, 75*, 258-267.

Ambach, W., Dummel, S., Lüer, T., Vaitl, D. (2011). Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. *International Journal of Psychophysiology*, *81*, 275-282.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavioral Research Methods*, *37*(3), 379-384.

Ben- Shakhar, G., Elaad, E. (2003). The validity of psychophysiological detection of deception with the Guilty Knowledge Test: a meta-analytic review. *Journal of Applied Psychology, 88*(1), 131–151.

Carmel, D., Dayan, E., Naveh, A., Raveh, O., Ben- Shakhar, G., (2003). Estimating the validity of the Guilty Knowledge Test from simulated experiments: the external validity of mock crime studies. *Journal of Experimental Psychology: Applied*, *9*(4), 261–269.

Cutmore, T. R. H., Djakovic, T., Kebbell, M. R. and Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychphysiology*, *71*, 185-192, doi:10.1016/j.ijpsycho.2008.08.003.

Dietrich, A. B., Hu, X., Rosenfeld, J. P. (2014). The Effects of Sweep Numbers Per Average and Protocol Type on the Accuracy of the P300- Based Concealed Information Test. *Applied Psychophysiological Biofeedback*, *39*, 67-73,doi:10.1007/s10484-014-9244-y.

Elaad, E. (2000). Detection of deception. *Encyclopedia of Forensic Sciences*, *1*, 550-556.

Farwell, L. A. (2012). Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics, 6*, 115-154, doi: 10.1007/s11571-012-9192-2.

Farwell, L. A., Donchin, E. (1991). The truth will out: Interrogative polygraphy (‘lie detector’) with event-related brain potentials. *Psychophysiology, 28*, 531-547.

Farwell, L. A., Smith, S. S. (2001). Using brain MERMER testing to detect knowledge despite efforts to conceal. *Journal of Forensic Sciences*, *46*, 135–143.

Folstein, J. R., van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*, 152–170.

Gamer, M., Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology, 47*, 355–364, doi: 10.1111/j.1469-8986.2009.00933.x.

Ganis, G., Keenan, J. P. (2009). The cognitive neuroscience of deception. *Social Neuroscience*, *4*(6), 465- 472.

Hu, X., Hegeman, D., Landry, E., Rosenfeld, J. P. (2012). Increasing the numbers of irrelevant stimuli increases ability to detect countermeasures to the P300- based Complex Trial Protocol for concealed information detection. *Psychophysiology*, *49*, 85-95.

Hu, X., Pornpattananangkul, N., Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology,* 50(5), 454-464, http://dx.doi.org/10.1111/psyp.12018.

Hu, X., Wu, H., Fu, G. (2011). Temporal course of executive control when lying about self- and other- referential information: An ERP study. *Brain Research,* 1369, 149-157, doi: 10.1016/j.brainres.2010.10.106.

Jokinen, A., Santtila, P., Ravaja, N., Puttonen, S., (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. *International Journal of Psychophysiology, 62*(1), 175–184, doi:10.1016/j.ijpsycho.2006.04.004.

Karton, I., Kitt, A. B., Rutiku, R., Bachmann, T. (2014). *Does the critical stimulus in a psychophysiological concealed information test have a unique status in terms of the revealing ERP components? It depends.* Submitted manuscript.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal Of Applied Psychology*, *43*(6), 385-388.

Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin*, *86*(1), 47-53.

Matsuda, I., Nittono, H., Allen, J. J. B. (2013). Detection of concealed information by P3 and frontal EEG asymmetry. *Neuroscience letter*, *537*, 55-59.

Mertens, R., Allen, J. J. B. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios.

Psychophysiology, 45, 286-298, doi: 10.1111/j.1469-8986.2007.00615.x.

Nahari, G., Ben- Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology, 48*, 733-744, doi: 10.1111/j.1469-8986.2010.01148.x.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. *In: Memory Detection: Theory and Application of the Concealed Information Test* (Verschuere, B., Ben-Shakhar, G., Meijer, E., eds), pp. 63-89. Cambridge: Cambridge University Press.

Rosenfeld, J. P., Biroschak, J. R., Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*(3), 251-9.

Rosenfeld, J. P, Hu, X., Labkovsky, E., Meixner, J., Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology, 90*, 118–134.

Rosenfeld, J. P., Hu, X., Pederson, K. (2012). Deception awareness improves P300-based deception detection in cocealed information tests. *International Journal of Psychophysiology, 86*, 114-121, doi: 10.1016/j.ijpsycho.2012.06.007.

Rosenfeld, J. P. Soskins, M., Bosh, G., Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology, 41*, 205- 219, doi: 10.1111/j.1469-8986.2004.00158.x.

Seymour, T. L., Seifert, C. M., Shafto, M. G., Mosmann, A. L. (2000). Using response time measures to assess ‘guilty knowledge’. *Journal of Applied Psychology*, *85*(1), 30–37.

Sip, K. E., Carmel, D., Marchant, J. L., Li, J., Petrovic, P., Roepstorff, A., et al. (2013). When Pinocchios`s nose does not grow: belief regarding lie- detectability modulates production of deception. *Frontiers in Human Neuroscience*, 7, 1-11.

Verschuere, B., Rosenfeld, J.P, Winograd, M. R., Labkovsky, E., Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal and Criminological Psychology, 14*, 253-262.

Winograd, M. R., Rosenfeld, J. P. (2011). Mock crime application of the Complex Trial Protocol (CTP) P300- based concealed information test. *Psychophysiology*, *48*, 155-161, doi: 10.1111/j.1469-8986.2010.01054.x.

Woodman, G. F. (2010). A brief introduction to the use of event- related potentials in studies of perception and attention. *Attention, Perception & Psychophysics, 72*(8), 2031-2046.

Wu, H., Hu, X., Fu, G. (2009). Does willingness affect the N2- P3 effect of deceptive and honest responses? *Neuroscience Letters*, 467, 63-66.

6. ACKNOWNLEDGEMENTS

I thank my supervisor Talis Bachmann and co-supervisor Inga Karton for the support during the preparation of this study and thesis. I would also like to thank Renate Rutiku and Anu Einberg for their advice and help in different stages of this study.

Non-exclusive licence to reproduce thesis and make thesis public

I, Anni- Bessie Kitt (date of birth: 24.06.1989),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the university's web environment, including via the DSpace digital archives, as of **20.05.2016** until expiry of the term of validity of the copyright,

"Detecting quilty knowledge with EEG- based methods using mock crime scenario", supervised by Talis Bachmann and Inga Karton,

2. I am aware of the fact that the author retains these rights.

3. This is to certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tallinn, **20.05.2014**