한국심리학회지 : 법
The Korean Journal of Forensic Psychology
2020, Vol. 11, No. 1, 89-115

Revalidation of the Complex Trial Protocol using participant-oriented countermeasures*

Hyemin Kim　　Inuk Song　　Eunhee Chang　　Hyun Taek Kim†

Department of Psychology, Korea University

Traditional deception detection methods had examined the difference of one's autonomic physiological responses through asking crime-related and crime-unrelated questions. There has been a continuing controversy regarding the accuracy and validity of the test, and thus, many researchers were motivated to explore and develop alternative efficient methods of detection in which one of them is known as P300-based Complex Trial Protocol (CTP). The P300-based CTP detects deception through comparing the P300 amplitudes between probe and irrelevant stimuli and is known as a counterstrategy of countermeasures. However, many previous studies have used countermeasures created from Rosenfeld et al.'s work (2008). The present study initially conducted a survey asking open-ended questions about the countermeasure use to acquire participant-oriented countermeasures for the main experiment. Then, the study aimed to evaluate whether the CTP can accurately detect deception even in the use of survey-based countermeasures. We firstly selected a set of participant-oriented countermeasures through survey questions. Then, a total of 50 participants were divided into three groups (innocent, guilty, and countermeasures) and performed the CTP. Those assigned to the countermeasures group covertly performed mental countermeasures during the CTP. The results of P300 amplitude analysis revealed that the guilty group's P300 amplitude of probe stimuli was significantly larger than that of irrelevant stimuli. Countermeasures group also had a significantly larger P300 amplitude for probe stimuli compared to irrelevant stimuli, even in the use of countermeasures. The results of bootstrapped amplitude difference (BAD) showed a detection accuracy rate of 81.25%, 82.35%, 82.35% for the innocent, guilty, and countermeasures groups, respectively. These findings demonstrate that the CTP can obtain a high detection rate in participant-oriented countermeasures and suggest the potential use of the CTP in the field.

Key words : Deception detection, Complex Trial Protocol, P300, countermeasures

---

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A01009211).
이 논문은 제1저자 김혜민의 석사학위 논문 일부를 발췌 수정한 것임.

† 교신저자 : 김현택, 고려대학교 심리학과, 서울특별시 성북구 안암동 5가 고려대학교 구법학관 205호
Tel : 02-3290-2065, E-mail : neurolab@korea.ac.kr

Since Marston's (1917) deception detection study, there has been a consistent research effort to investigate psychophysiological detection of deception (e.g., Lykken, 1959, 1979; Reid, 1945). Traditionally, a lie detection has assumed that a distinct physiological response, such as increased galvanic skin response (Podlesny & Raskin, 1977), is associated with lying. The most well-known traditional method is the Comparison Question Test (CQT; Marino, Reid, & Inbau, 1977), which based its rationale on the variability in one's physiological response when answering crime-related and crime-unrelated question (Raskin, 1989). There has been a continuing controversy regarding the accuracy and validity of the CQT (e.g., Lykken, 1984; Patrick & Iacono, 1991; Saxe, Dougherty, & Cross, 1985). Lykken (1979) argued that when the polygraph charts were scored blindly, the accuracy of the CQT decreased to 64% against a chance expectation of 50%. Moreover, Lykken criticized the high false positive error rate of the CQT, biasing against innocent people.

A new alternative detection method called the Guilty Knowledge Test (GKT), also known as the Concealed Information Test (CIT), was proposed by Lykken (1959) to decrease the error rate of a lie detection. The GKT detects a "guilty knowledge" only a guilty person can recognize by observing one's autonomic responses. A crime-related information a suspect might have used it or seen it in a crime scene, such as a weapon or stolen object, defines "guilty knowledge". An item that contains a guilty knowledge is called a *probe* (e.g., gun, knife, diamond ring, etc.). A probe stimulus is interspersed with other items, called an *irrelevant*, that are similar but not related to crime. Number of probe and irrelevant stimuli are presented to a suspect in random order. To keep their focus on the test, they are required to stimulate different response (e.g., pressing a different button) to one of the irrelevant stimuli, and this stimulus is called a *target*. The investigator, then, measures the suspect's autonomic physiological response to determine whether the suspect elicits a significantly different signal when the probe appears compared to the appearance of the irrelevants. Autonomic physiological responses had been used in the past (Ginton, Daie, Elaad, & Ben-Shakhar, 1982; Horowitz, Kircher, Honts, Raskin, 1997; Podlesny & Raskin, 1977), yet this measurement was found to be unstable as the emotional state of an examinee (e.g., fear, anxiety, etc.) might involuntarily affect their autonomic physiological response (Ben-Shakhar, Lieblich, & Bar-Hillel, 1982). For that reason, some researchers turned their attention to an event-related potential (ERP) measure, especially the P300 component that is known to reflect one's cognitive evaluation of a given stimulus (Sutton, Tueting, Zubin, & John, 1967).

P300 amplitude and its relation to

deception detection

First and foremost, it is critical to understand the functional basis of P300. In the late 1900s, Donchin and his colleagues explored the function of the P300 component in an attempt to develop a measure of one's "mental workload" (Donchin & Coles, 1988; Gopher, & Donchin, 1986; Polich, 2007). According to their work, two factors control the amplitude of P300: subjective probability and task relevance (Donchin & Coles, 1988). For example, if a thief was shown a set of pictures in which he unexpectedly finds a picture of the ring he stole, then his subjective probability for the ring picture would increase, thus eliciting larger P300 component, whereas others who are unaware of the theft incident would have low subjective probability and would not elicit large P300 component. Task relevance is the extent to which one allocates their attention to an assigned task. When participants did not pay attention to the given stimuli, P300 was not elicited (Duncan-Johnson & Donchin, 1977). This nature of P300 component can be utilized in the field of deception detection, especially in context of finding a guilty knowledge that only a criminal can have. The CIT was used with measuring ERPs of subjects, referred as P300-based CIT. The P300-based CIT has been mainly studied by Rosenfeld and his colleagues and its detection accuracy was reported to be near 90% (Allen, Iacono, & Danielson, 1992; Hu & Rosenfeld, 2012; Rosenfeld, 2002).

Countermeasures and counterstrategy

Despite continuous improvements of lie detection methods, some limitations still remain. One of the limitations that has been repeatedly reported in the literature is the vulnerability to countermeasures (Ben-Shakhar & Dolev, 1996; Honts, Raskin, & Kircher, 1994; Meixner & Rosenfeld, 2010; Mertens & Allen, 2008; National Research Council, 2003; Rosenfeld, Soskins, Bosh, & Ryan, 2004; Sasaki, Hira, & Matsuda, 2001; Sokolovsky, Rothenberg, Labkovsky, Meixner, & Rosenfeld, 2011). A countermeasure refers to a tactic or a strategy that an examinee intentionally and covertly use to cheat on a lie detector with two major categories: mental countermeasures (e.g., production of emotional imagery) and physical countermeasures (e.g., breathe control, biting one's tongue) (National Research Council, 2003). Rosenfeld et al.'s study (2004) showed that assigning different concealed responses (countermeasures) to each irrelevant item and participants covertly executing the countermeasures - for example, pressing the left forefinger on the leg or imagining the experimenter slapping the participant in the face - is an effective way of defeating the P300-based CIT. These countermeasures make the corresponding irrelevant items relevant during the task. Consequently, the oddball effect is diminished, resulting in reduced differences between the P300 amplitudes of the probe and irrelevant items.

The CIT requires two concurrent tasks, explicitly discriminating a target stimulus and implicitly recognizing a probe stimulus. According to several studies conducted by Isreal and his colleagues, two concurrently executing tasks can divert one's attention and thereupon deplete the processing resources and the amplitude of P300 (Isreal, Chesney, Wickens, & Donchin, 1980). When subjects were instructed to perform a main task that varies cognitive demands and concurrently performed a secondary task of mentally counting target oddball stimuli, a reciprocal relationship between the cognitive difficulty of the main task and the P300 amplitude from the oddball (secondary) task was found. The more demanding the main task, the smaller the P300 amplitude (Isreal et al., 1980; Kramer, Wickens, & Donchin, 1985; Kramer, Wickens, Vanasse, Heffley, & Donchin, 1981).

Rosenfeld and his colleagues created the The Complex Trial Protocol (CTP) to counter the use of countermeasures through splitting the tasks into two parts, preventing one's attention to divert into two concurrent tasks (Rosenfeld et al., 2008). He also discussed that P300 component and a reaction time (RT) can be the indicators of a lie detection. The procedure of the CTP is as follows: the first task requires pressing the same button for both probe and irrelevant stimuli to make respondents solely concentrate on an implicit response. It is then followed by the second task of discriminating (pressing a different button) a target stimulus among nontarget stimuli, which involves an explicit response. Rosenfeld et al.'s work (2008) demonstrated that in this first task the guilty and those who used countermeasures had larger P300 amplitude than that of the innocent for the probe stimulus. And the second task is only to check if participants are focused on the task they are given. If they conscientiously execute the task, a target stimulus will have larger P300 amplitude than nontarget stimuli's. Moreover, RT of the participants who used countermeasures for irrelevant stimuli was longer than probe stimulus, whereas RT of the guilty for probe stimulus was longer than irrelevant stimuli (Rosenfeld et al., 2008). It is because those who use countermeasures take longer time to think and use countermeasures before pressing the button for irrelevant stimuli compared to recognizing the probe stimulus. On the contrary, the guilty do not have additional work besides recognizing the probe stimulus. The CTP, therefore, was created as a counterstrategy of countermeasures.

Purpose of the present study

Although a myriad of studies have examined the effect of the CTP on detecting deception (e.g., Hu & Rosenfeld, 2012; Meixner, Haynes, Winograd, Brown, & Rosenfeld, 2009; Rosenfeld, Tang, Meixner, Winograd, & Labkovsky, 2009; Rosenfeld, Ward, Thai, & Labkovsky, 2015; Sokolovsky et al., 2011), only one study (Winograd & Rosenfeld, 2011) has applied a mock crime scenario and the rest depended on using autobiographical plot, which is known to elicit stronger and larger P300 amplitude (Rosenfeld, Biroschak, & Furedy, 2006). There is a limited amount of research using mock crime scenario as the contents for CTP. Furthermore, because Winograd and Rosenfeld (2011) mainly used physical countermeasures to defeat against CTP, assigning all mental countermeasures with a mock crime scenario is necessary since they are known as the least detectable and the fastest to use (Lukács et al., 2016; Rosenfeld & Labkovsky, 2010). Using all mental countermeasures in CTP with mock crime scenario has not yet investigated.

Moreover, none of the related studies have surveyed participants in regard to countermeasures but rather assigned countermeasures that were used in Rosenfeld et al.'s work (2004). We thus utilized a survey, explaining lie detection and asking open-ended questions about the types of countermeasures they would use, to embrace the participant-oriented approach as we later applied them into the main experiment. The participant-oriented approach, in this paper, means that we emphasized the views of participants and reflected their opinions to the experiment. We expected using a survey-based countermeasures would put the CTP into greater challenge since the survey results can demonstrate what people would actually do during the investigation to deceive the task. If the result indicates the CTP successfully distinguishing the countermeasure use, then it may imply the possibility of the CTP having more flexible functionality.

Also, only Rosenfeld and his laboratory conducted most studies related to the CTP, except for Lukács et al.'s work (2016). The importance of replication study is already mentioned in the review (Ben-Shakhar, 2012). Therefore, the present study aims to examine and revalidate the functionality of CTP using participant-oriented mental countermeasures in a mock crime scenario.

We divided participants into three groups: Guilty (G), Countermeasure (CM), and Innocent (IC). Participants in the G and CM group were asked to commit a mock crime and those in the IC group executed non crime related mission. Based on the logic that the CTP was made to divide two concurrent tasks, thus enhancing one's attention (Rosenfeld et al., 2008), and the oddball effect of guilty knowledge, we hypothesized that the CTP would successfully detect deception. To be more specific, we expected, between irrelevant and probe stimuli, longer irrelevant's RT for CM and longer probe's RT for G, which replicates the RT result of Rosenfeld et al.'s study (2008). Also, we expected the CTP successfully detecting and discriminating between G and IC just like the CIT does through analyzing P300 component and detecting CM even when participants try concealing their guilt by using countermeasures.

Survey

Prior to the main experiment, the survey was conducted using Amazon's Mechanical Turk for the purpose of selecting participant-oriented countermeasures. The main question was "If you were a criminal, what kind of countermeasure are you going to use in order to defeat the CIT?" All respondents read a description explaining the definition of countermeasures and its relation to lie detection and took a quiz before answering the survey. They received $0.65 for their participation.

A total of 75 respondents (33 female) answered the survey. The mean age was 35 years. The responses were categorized into two types of countermeasures: physical and mental. The detailed results of the survey are shown in Table 1. Because physical countermeasures can be more noticeable and detectable than mental countermeasures and also because we wanted to evaluate the CTP in difficult situation, physical countermeasures were excluded from the survey data. Also, we followed Rosenfeld et al.'s (2008) means based on the responses in the mental countermeasures category.

Table 1. Survey results

Types of countermeasures	Number of respondents
Physical countermeasures	
Stay calm; meditate	8
Control breathing	6
Control heart rate	4
Control muscle contractions	4
Inflict pain with something sharp	1
Mental countermeasures	
Change irrelevant into relevant (connecting irrelevant stimuli to the crime)	21
Think about something else during the irrelevant items	10
Create false memory about crime	6
Think of the crime the entire time	5
Distract with other responses	4
Change probe into irrelevant	3
Make vague answers	3

With the aims of (1) adopting participants' prevailing opinion into the countermeasures and (2) simplifying the use of countermeasures, we constructed mental countermeasures based on the survey data (See Methods).

Methods

Participants

50 participants (26 female) were recruited through online advertisements. All participants were right-handed and had normal or corrected normal vision. Their ages were 19 - 27 years. The participants were randomly assigned to one of the three groups: IC (n = 16), G (n = 17), or CM (n = 17). An institutional board review was obtained from the Korea University (KUIRB-2018-0050-01).

Procedure

The experiment was divided into two phases: (1) an initial phase, where the assistant guided the participants to complete questionnaires, mock crime, and a supplementary task, and (2) a second phase, where the investigator proceeded by conducting a short interview and the CTP. A diagram of the experimental procedure is shown in Figure 1.

In their arrival, all participants were guided to a designated room and the assistant explained the overall procedure of the experiment. Monetary compensation (40,000 won) was given to the participants. For the sake of engagement, however, we insisted they must be judged innocent in order to receive full compensation and they will only receive the half if we find them guilty. In other words, the G group pretended to be innocent trying their best concealing the truth, the CM

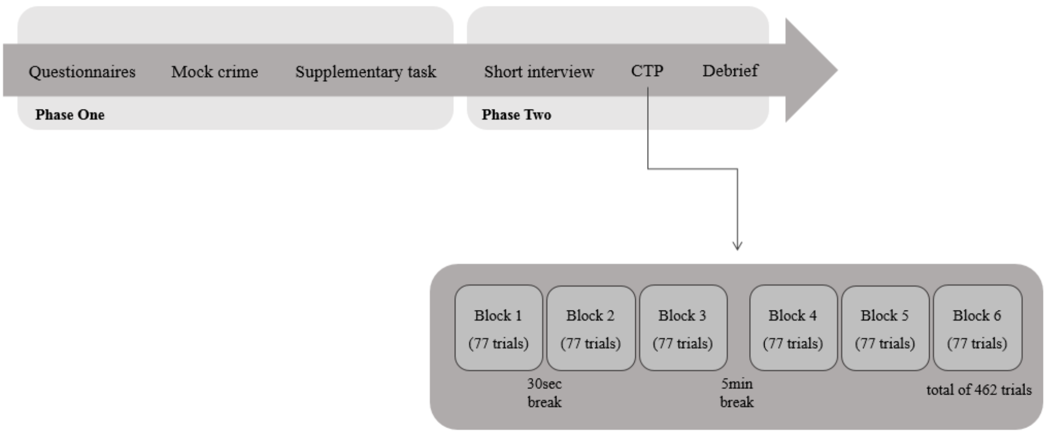

Figure 1. Experimental procedure

group pretended be innocent and additionally used countermeasures, without being caught, to deceive the investigator, and the IC group behaved honestly. Actually, all participants received full compensation regardless of the verdict. The written informed consent of participation was obtained prior to the experiment. The participants, then, filled in the questionnaires and picked one of the two envelopes containing instruction sheet of either a guilty mission or an innocent mission. The mission, in fact, was individually assigned to each participant beforehand, but they were unaware of the assignment. Participants read the instruction sheet and were asked to summarize the procedure making sure they fully understood the mission.

The general scenario was that the student (participant) was to go inside Professor Kim's office, find a gift and packing materials, and wrap it with a ribbon. Participants in the G and CM groups participated in a mock crime. They were instructed to turn off a CCTV camera located in Professor Kim's office, steal jewelry from a gift box while wrapping, and turn on the CCTV camera before leaving. Half of the participants stole a ring and other half stole a watch to counterbalance the use of probe. Participants in the innocent group did not participate in a mock crime, but went inside the office and wrapped a gift. They were instructed not to open a gift box.

After the completion of mission, participants came back to the room to work on a supplementary task. Participants in CM group were told that they will be learning a trick that can help them pass a lie detection without getting caught (countermeasure). The information sheet elucidating (1) what the CTP is, (2) how P300 of ERP can differentiate between the G and IC, and (3) the definition of countermeasure is and how it can affect a lie detection were passed to the participants. Basically, they learned basic logic behind EEG-based lie detection and countermeasures. As a mental countermeasure, they were specifically told to think of very intense, arousing, or brutal thought using the irrelevant stimulus shown on the screen without getting caught by the interrogator. For example, if the irrelevant stimulus on the screen is "fountain pen", they might think of stabbing a person with a fountain pen. They were instructed to use this mental countermeasure for all irrelevant stimuli. This type of mental countermeasure used in the present was selected based on the survey results. We highly emphasized the fact that the interrogator should not notice them using countermeasure. They were also informed to think of the thought only when the irrelevant stimulus is present and to press the button after thinking. They were quizzed so that we knew they thoroughly understood what they had to do in the interrogation. Then, they practiced the CTP until they could use countermeasure on time. We ensured they could proficiently use countermeasure during the actual test.

Meanwhile, participants in the G and IC group read an article unrelated to the experiment for a mere reason of reaching similar workload with the CM group. The article discussed the efficacy of almonds and Brazil nuts as a nutritious food supplement. The participants were required to answered questions related to the article. They practiced a somewhat similar CTP with all stimuli switched to the types of the nuts and vitamins. They were told the reason of doing a supplementary task was to distract them away from mock crime before going into the interrogation. We strongly warned all participants that the interrogator does not know whether they are guilty or innocent and they have to be judged innocent based on the CTP results.

The assistant and participant walked to the interrogation room located in the different building. Upon the arrival of the participant, the interrogator inquired about the health condition that may affect the recording of EEG, asked the participant's whereabouts and actions before coming to the interrogation room, and requested to write an affidavit. After that, the participant was seated in front of the 17inch CRT display with the electroencephalography (EEG) cap placed on the head for the electrode preparation, which took approximately 20 to 30 minutes. The main task, the CTP, was conducted while EEG and electrocardiogram (ECG) signals were being recorded. A total of six blocks of trials (Fig.1) were given to all participants. After the CTP, participants in the CM group were asked to write the specific countermeasures they used for each irrelevant item. All participants were debriefed concerning the purposes of the experiment and thanked for their participation.

Complex Trial Protocol

The CTP structure used in Rosenfeld et al. (2008) was replicated with minor modifications. The words "ring" and "watch" were alternately used as a probe stimulus, and other words, "wallet", "necklace", "perfume", "fountain pen", "belt", and either "watch" or "ring" depending on a used probe, were the six irrelevant stimuli. Figure 2 shows an example of a trial in the CTP. Each trial started with a fixation cross in the center of the screen for 500ms, followed by a randomly varying delay of 1000, 1150, 1300, or 1450ms. After the delay, stimulus one (S1), either the probe or one of six irrelevants, was presented for 500ms and participants were instructed to press the left button on the Cedrus response box (RB-740) with their left index finger as soon as they saw the first stimulus. They pressed the left button as the means of seeing the words on the screen. A randomly varying time of 1400, 1550, 1700, or 1850ms was provided for the response one. Then, stimulus two (S2), either the target or one of four nontargets, was shown for 500ms. "11111" was a target stimulus and participants were

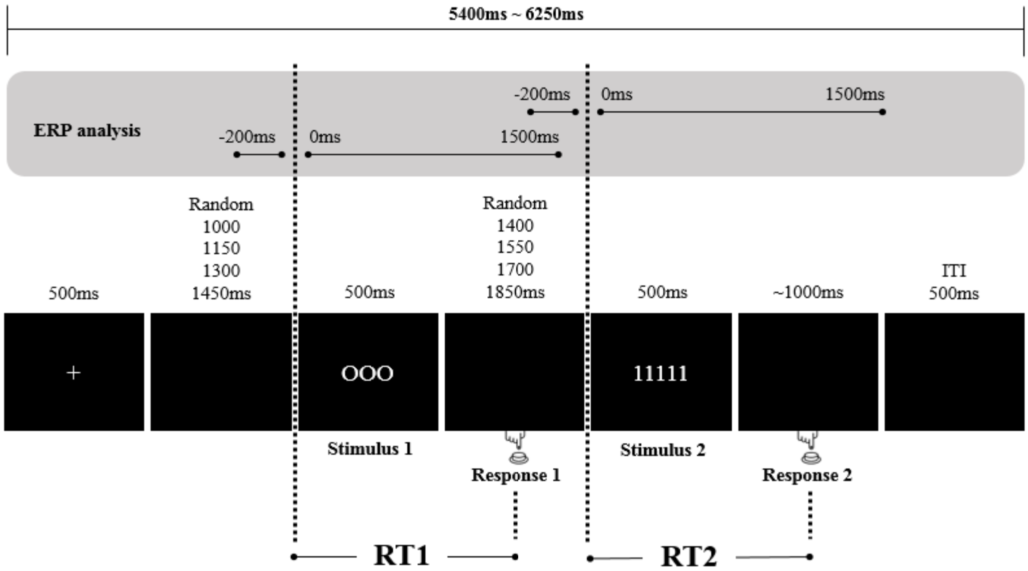

Figure 2. Example of a trial in the CTP

instructed to press the right button with their right index finger. "2222", "3333", "4444", or "5555" were the four nontarget stimuli and participants pressed the left button. A maximum of 1000ms was given for the response two. The inter-trial interval (ITI) was 500ms. The CTP was composed of six blocks (77 trials each) and a total of 462 trials. The visual angle varied between 2~4° for each stimulus. The E-Prime 2.0 software (Psychology Software Tools, Inc., Sharpsburg, USA) was used to construct the CTP trials.

Data Acquisition

EEG was recorded with SynAmps RT NeuroScan 64-channel EEG system with Quik-Cap 64 channels (NeuroScan Compumedics, Charlotte, NC, USA) with sampling rate of 1000Hz and impedance lower than 5kΩ. EEG electrodes were referenced to linked-earlobes. A grounded electrode was located in between Fpz and Fz electrodes. Four monopolar electrooculogram (EOG) electrodes were recorded (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015): two electrodes were vertically placed on the upper and lower side of the left eye (UVEO, LVEO) and other two electrodes were placed on the temples lateral to the eyes (LHEO, RHEO).

Moreover, ECG was recorded simultaneously by MP150 (BIOPAC Systems, Santa Barbara, CA, USA) with in-house trigger distribution system to make trigger marks simultaneously at a sampling rate of 1000Hz. ECG signals were used to eliminate the fluctuations caused by the heartbeats (Park, Correia, Ducorps, & Tallon-Baudry, 2014).

Questionnaires

A total of eight self-reported questionnaires were administered to assess participant's personality traits that may influence the results of the experiment: Handedness scale, Behavioral Inhibition System/Behavioral Activation System (BIS/BAS scale; Carver & White, 1994), Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), Self-Monitoring Scale (Snyder & Gangestad, 1986), Machiavellianism test (Mach-IV; Christie & Geis, 1970), Risk-Taking Questionnaire (RTQ; Knowles, Cutter, Walsh, & Casey, 1973), Beck Depression Inventory (BDI; Beck, Emery, & Greenberg, 2005), Beck Anxiety Inventory (BAI; Loeb, Beck, Diggory, & Tuthill, 1967). The BIS/BAS scales include four subscales: BIS, BAS-Drive, BAS-Fun Seeking, and BAS-Reward Responsiveness. PANAS scales consist of one scale measuring one's positive emotion and the other measuring the negative emotion. BIS/BAS scores associate with EEG asymmetry, PANAS, BDI, and BAI associate participants' emotions during the participation, and Mach-IV and RTQ relate to ability of lying (Song, Kim, Lee, Chang, & Kim, 2019). We replicated the selection of questionnaires from Song et al.'s study (2019). The results of all questionnaires were expected to have no significant differences among three groups since the difference might be considered as confounding factor.

Behavioral measures

The reaction times to the first stimulus (S1) were measured to evaluate whether it can used as an indicator of countermeasure use. The reaction times for target or nontarget were measured as well. See Figure 2 for the measured range of the reaction times. Furthermore, accuracies were calculated to ensure that the participants focused on the task.

Data processing

EEGLAB toolbox (Delorme & Makeig, 2004) in MATLAB (MathWorks Inc., Natick, USA) was used to preprocess EEG and ECG signals. ECG signals were aligned with EEG data by the first stimulus trigger points. The aligned data was downsampled to 250Hz. Winkler and colleagues (Winkler, Debener, Muller, & Tangermann, 2015) emphasized that 1Hz high-pass filtering produced significantly improved ICA calculation results. The data was, on that account, high-pass filtered with two cutoff frequencies using FIR filter, creating two distinct data with different high-pass filter: 0.5Hz cutoff frequencies for artifact rejection and independent component analysis (ICA; Hyvärinen & Oja, 2000) and 0.1Hz cutoff frequencies for artifact correction.

The data with 0.5Hz cutoff frequencies was low-pass filtered with 100Hz cutoff frequencies. 60Hz line noise was removed using PREP plugin with multitaper method (Bigdely-Shamlo et al., 2015). The entire recording was visually inspected to reject data and channels with detected artifacts (such as muscular activity) and the rejected data were interpolated. Using Adaptive Mixture Independent Component Analysis (AMICA; Palmer, Kreutz-Delgado, & Makeig, 2012), the ICA was calculated to separate independent subcomponents of the EEG.

The data with 0.1Hz cutoff frequencies was also low-pass filtered with 100Hz cutoff frequencies. The artifact rejection information from 0.5Hz cutoff frequencies data was imported and applied to the 0.1Hz cutoff frequencies data. Artifact correction was performed using Semi-Automated Selection of Independent Components of the electroencephalogram for Artifact correction (SASICA; Chaumon, Bishop, & Busch, 2015), which provides a list of artifact ICA components as a recommendation. Epochs starting at 300ms before stimulus onset and ending at 1500ms after stimulus onset were extracted (from -300ms to 1500ms, Fig.2) and were once again visually inspected for artifact removal.

ERP analysis: P300

The epochs were low-pass filtered with 30Hz cutoff frequencies, re-epoched from -200ms to 1500ms, and averaged. A baseline correction was applied from -200ms to 0ms. The data at Pz were used for the statistical analyses. Peak-to-peak measurement was chosen to calculate P300 amplitude (Rosenfeld et al., 2008; Soskins, Rosenfeld, & Niendam, 2001). Peak-to-peak method measures the difference between the highest amplitude within a specific window and the subsequent lowest amplitude. The epochs were passed through a moving average filter with 100ms time window, shifting 4ms at a time. The highest point was searched in the range of 300ms - 800ms. The range of right after the highest point (latency) - 1500ms was used for the lowest point. The difference value is used as the peak-to-peak P300 amplitude.

Bootstrap analysis: Individual detection

The bootstrap method (Wasserman & Bockenholt, 1989) was used on the Pz to classify participants as guilty or innocent. The procedure is as followed: (1) Individual probe trials were sampled randomly, with replacement, from all probe single trials, and averaged into one re-sampled epoch. (2) Individual irrelevant trials were sampled randomly, with replacement, from all irrelevant single trials, and averaged into one re-sampled epoch. (3) The P300 peak-to-peak amplitude of probe trials and that of irrelevant trials were compared. A value of 1 indicates the amplitude of probe is greater than that of irrelevant and a value of 0 indicates the amplitude of probe is less than that of irrelevant. (4) This procedure was repeated 1000 times (1000 iterations), successively adding all values. The final result was a value between 0 and 1000. If the final value is greater than 900 (cutoff of 90%; Farwell & Donchin, 1991), then the participant is classified as guilty. If the final value is less than 900, then the participant is classified as innocent. In addition, a receiver operating characteristic curve (ROC curve) was calculated to evaluate the diagnostic accuracy of individual classification. MedCalc (MedCalc Software, Mariakerke, Belgium; Schoonjans, Zalata, Depuydt, & Comhaire, 1995) was used to compare the areas under curve of ROC between the guilty and CM groups.

Results

Questionnaires

A one-way ANOVA was performed to assess whether the three groups differed on the scores. The results showed no significant differences between the groups for any of the questionnaires. The mean scores and standard deviations for each questionnaire are listed in Table 2.

Table 2. Questionnaire score results

	IC	G	CM	*F*	*df*	*p*
Age	21.9 (2.9)	22.5 (3)	22.3 (2.1)	0.25	(2,47)	0.78
Handedness Scale	31.4 (8.2)	28 (7.4)	30.5 (7.5)	0.88	(2,47)	0.42
BIS	18.9 (1.8)	19.9 (2)	19.9 (2.7)	1.23	(2,47)	0.3
BAS-Reward	17.8 (2.3)	16.5 (1.9)	17.4 (2.1)	1.86	(2,47)	0.17
BAS-Drive	12.2 (1.7)	11 (2.5)	11.5 (1.8)	1.52	(2,47)	0.23
BAS-Fun Seeking	11.6 (2.4)	11.8 (1.3)	12.5 (2.4)	0.72	(2,47)	0.49
PANAS-positive	23.6 (6.5)	22.9 (6.1)	22.8 (5.8)	0.1	(2,47)	0.91
PANAS-negative	14.9 (5.8)	15.9 (3.5)	12.9 (2.5)	2.62	(2,47)	0.11
Self-Monitor	8.5 (2.6)	8.6 (2.6)	9.1 (1.9)	0.27	(2,47)	0.76
Mach-IV	58.8 (6.2)	58.3 (5.1)	58.6 (7.6)	0.02	(2,47)	0.98
Risk Taking	64.4 (6.2)	63.4 (8.6)	61.5 (6.3)	0.69	(2,47)	0.51
BAI	6.5 (6.5)	9.1 (7.1)	7.6 (8.6)	0.54	(2,47)	0.59
BDI	5.5 (3)	7.4 (6.9)	5.8 (5.1)	0.58	(2,47)	0.57

Behavioral measures

Table 3 presents the mean reaction times and accuracies with standard deviations for all conditions, and Table 4 presents the statistical results.

RT: First task (probe vs. Iall)

Six irrelevant reaction times were combined and averaged (Iall) for the analysis. A Greenhouse-Geisser correction was used where the assumption of sphericity has been violated. A repeated measures ANOVA across all groups showed no significant reaction time differences within individual irrelevants, $F(7.393,177.427) = 0.687$, $p = 0.691$. A 3 x 2 mixed-model ANOVA with group (IC, G, CM) as between factor and stimulus type (probe, Iall) as within factor was performed. A significant main effect of group, $F(2,47) = 9.831$, $p < 0.001$, $\eta_p^2 = 0.291$, was found, but the main effect of stimulus type, $F(1,49) = 3.362$, $p = 0.073$, $\eta_p^2 = 0.063$, was not significant. Bonferroni post-hoc analysis revealed that the participants' reaction times were significantly slower for the CM group than for the IC ($p < 0.01$) and G group ($p < 0.01$). There was a significant interaction between group and stimulus type, $F(2,47) = 7.836$, $p = 0.001$, $\eta_p^2 = 0.291$. A paired sample t-test was used to compare reaction times of the probe and Iall within each group. The results showed a significant difference

Table 3. Mean reaction times (SD) and accuracies (SD) for different types of stimuli

	IC	G	CM
Reaction Times (ms)			
First task			
Probe	482.45 (193.4)	491.39 (178.36)	696.78 (188.32)
Iall	484.82 (197.79)	470.97 (182.38)	796.63 (278.95)
Second task			
Target	549.96 141.59)	581.96 (149.75)	641.96 (147.33)
Nall	535.88 (185.15)	524.93 (175.47)	608.32 (174.12)
Accuracies (%)			
First task			
Probe	99.36 (1.45)	99.37 (0.94)	99.2 (1.62)
Iall	99.6 (0.5)	99.76 (0.32)	99.66 (0.38)
Second task			
Target	96.31 (6.17)	94.52 (5.48)	94.11 (5.54)
Nall	99.14 (0.87)	99.37 (0.57)	98.7 (1.92)

Table 4. Reaction times results of repeated measures ANOVA and post analyses with t test (* $p<0.05$, ** $p<0.01$, *** $p<0.001$)

	F	*df*	*p*	η_p^2
Group	1.32	(2,47)	0.277	0.052
Stimulus type	18.075	(1,49)	0.000***	0.266
Group x Stimulus type	2.415	(2,47)	0.1	0.091
	t	*df*	*p*	
IC	1.08	15	0.296	
G	3.695	16	0.002**	
CM	2.608	16	0.019**	

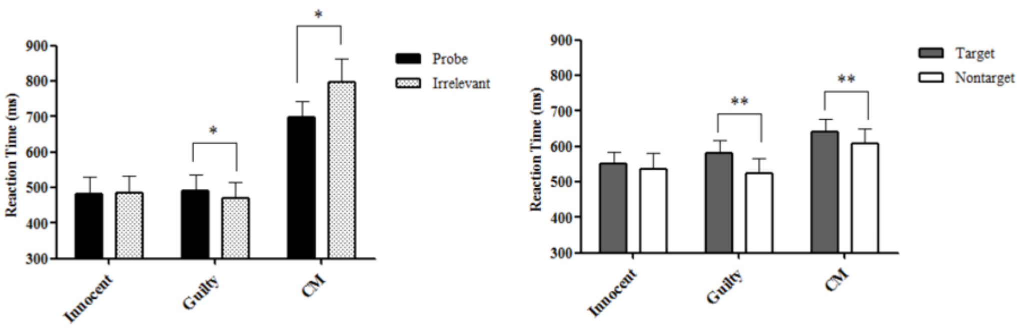

Figure 3. Between-group comparisons of mean reaction times. Reaction times of the first section (left), and reaction times of the second section (right)

for the G ($t(16) = 2.18$, $p = 0.045$) and CM groups ($t(16) = -2.613$, $p < 0.05$) only (See Fig.3). The participants in the G group had a slower response time to probe words than to irrelevant words, whereas the participants in the CM group responded faster to probe words compared to irrelevant words. No difference was observed in the IC group ($t(15) = -0.626$, $p = 0.54$). The results successfully replicated Rosenfeld et al.'s behavioral measures (2008).

RT: Second task (target vs. Nall)

Nall refers to four nontarget reaction times that are combined and averaged for the analysis. A repeated measures ANOVA across all groups showed no significant reaction time differences within individual nontargets, $F(6,144) = 0.984$, $p = 0.438$. A 3 x 2 mixed-model ANOVA (group: IC, G, CM x stimulus type: target, Nall) was conducted. A significant main effect of stimulus type was revealed, $F(1,50) = 18.075$, $p < 0.001$, $\eta_p^2 = 0.266$, but the main effect of group (F(2,47) = 1.32, p = 0.277, η_p^2 = 0.052) and its interaction with stimulus type (F(2,47) = 2.415, p = 0.1, η_p^2 = 0.091) were not significant. In other words, the participants' responses to target stimuli took significantly longer than to nontarget stimuli. Reaction times of the target and Nall within each group was compared using a paired sample t-test. A significant difference was found for the G (t(16) = 3.695, p < 0.01) and CM groups (t(16) = 2.608, p < 0.05) (See Fig.3). The participants in the G and CM groups responded faster to the nontarget stimuli compared to the target stimuli. There was no significant difference in the IC group (t(16) = 1.08, p = 0.296). The reaction times for target and nontarget stimuli do not specifically indicate detection of deception but clarify participants' attention towards the test. Although the IC group did not have any significant difference, all three groups had longer reaction time for target stimulus than for nontarget stimuli.

Accuracies

In the first section, the accuracy was close to perfect in all conditions, with no significant main effects or interactions, since all stimuli required the same response. Regarding the second section, a significant main effect for stimulus type was observed (F(1,49) = 28.108, p < 0.001), revealing that participants had lower accuracies for target stimulus ("11111") than for nontarget stimuli (94.98 ± 5.7% vs. 99.07 ± 1.26%). Because a target stimulus was the only stimulus assigned to different button, it is plausible to have lower accuracy compared to other stimuli.

ERP: P300

The mean P300 amplitudes of all stimuli (with SD) are shown in Table 5 and statistical results are shown in Table 6. Figure 5 illustrates the grand average of P300 waveforms for the IC, G, and CM groups, respectively.

Table 5. Means (SD) of P300 amplitudes at Pz

	IC	G	CM
P300 amplitude (peak-to-peak)			
First task			
Probe	3.95 (2.14)	6.79 (2.98)	7.95 (4.87)
Iall	3.65 (2.05)	4.2 (2.19)	4.11 (2.2)
Second task			
Target	12.25(6.25)	9.9 (3.6)	8.36 (4.23)
Nall	5.03(2.33)	5.26 (2.87)	4.51 (2.06)

Table 6. P300 amplitudes results of repeated measures ANOVA and post analyses with t test (* p<0.05, ** p<0.01, *** p<0.001)

	F	*df*	*p*	η_p^2
Group	3.071	(2,47)	0.056	0.116
Stimulus type	29.724	(1,49)	0.000***	0.378
Group x Stimulus type	7.462	(2,47)	0.002**	0.241
	t	*df*	*p*	
IC	0.766	15	0.456	
G	6.178	16	0.000***	
CM	4.062	16	0.001**	

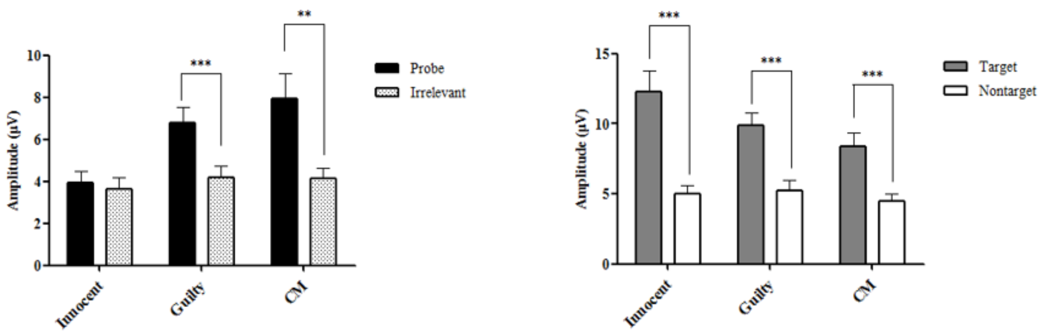

Figure 4. Between-group comparisons of mean P300 amplitudes. Mean P300 amplitudes (probe vs. Iall) at Pz of the three groups (left), and mean P300 amplitudes (target vs. Nall) at Pz of the three groups (right)

First task (probe vs. Iall)

P300 amplitude at Pz was calculated with peak-to-peak method. A mixed-measure ANOVA was used for the analysis of the P300 amplitude with group (IC, G, CM) as between-factor and stimulus type (probe, Iall) as within-factor. Significant interaction of group and stimulus type was observed ($F(2,47) = 7.462$, $p < 0.01$, $\eta_p^2 = 0.241$). The main effect of stimulus type ($F(1,49) = 29.724$, $p < 0.001$, $\eta_p^2 = 0.378$) was found, revealing that probe stimuli elicited significantly larger P300 amplitudes than Iall stimuli, but the main effect of group ($F(2,47) = 3.071$, $p = 0.056$, $\eta_p^2 = 0.116$) was not found. A paired-sample t test indicated the amplitude of probe was larger than that of Iall for the G ($t(16) = 6.178$, $p < 0.001$) and CM groups ($t(16) = 4.062$, $p < 0.01$). The IC group did not have different amplitudes ($t(15) = 0.766$, $p = 0.456$) (See Fig.4).

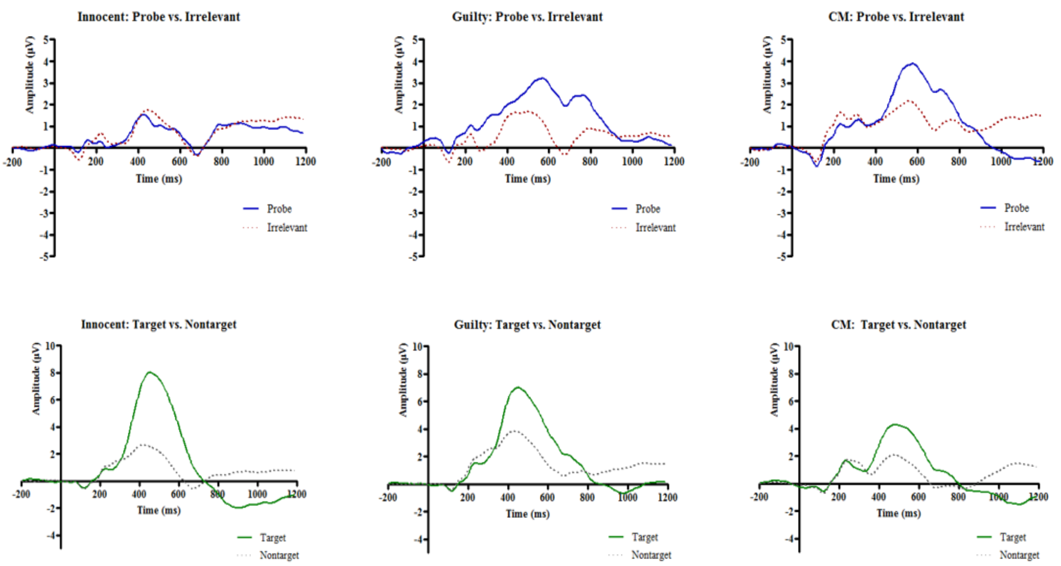

Figure 5. Grand average of P300 waveforms for the three groups at Pz

Second task (target vs. Nall)

A mixed-measure ANOVA showed a significant interaction between group (IC, G, CM) and stimulus type (target, Nall), $F(2,47) = 3.783$, $p < 0.05$, $\eta_p^2 = 0.139$. The main effect of stimulus type was significant ($F(1,49) = 86.948$, $p < 0.001$, $\eta_p^2 = 0.64$), but not for the group ($F(2,47) = 1.983$, $p = 0.149$, $\eta_p^2 = 0.078$), indicating that target stimuli evoked significantly larger amplitudes than Nall stimuli. All three groups had higher probe amplitude compared to Iall (innocent: $t(15) = 5.671$, $p <$

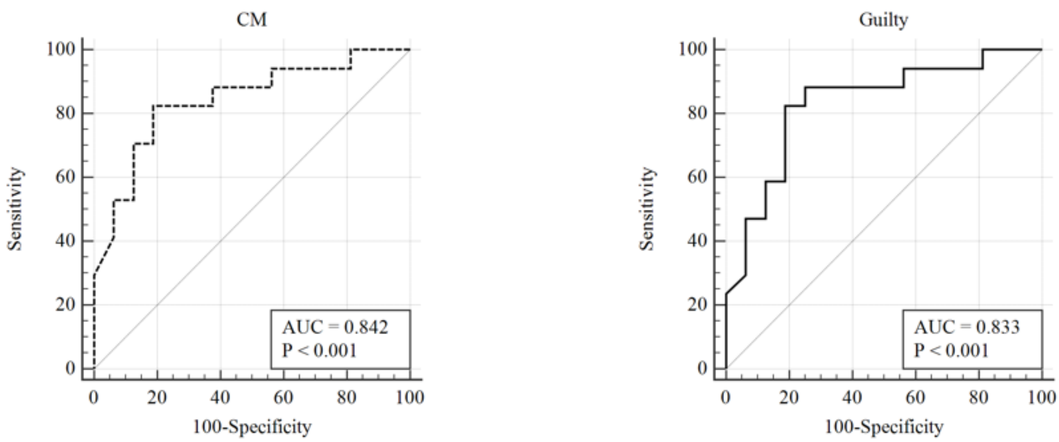

Figure 6. ROC curve of guilty and CM groups

Table 7. Bootstrap results for each participant

Subject	Probe vs. Iall			Target vs. Nall		
	IC	G	CM	IC	G	CM
1	**G (962)**	G (996)	**I (756)**	983	991	956
2	I (790)	**I (777)**	G (996)	976	999	989
3	**G (999)**	G (911)	G (998)	1000	997	1000
4	I (626)	G (994)	G (999)	1000	1000	901
5	I (613)	G (999)	G (916)	1000	1000	987
6	I (896)	G (956)	G (1000)	1000	996	1000
7	I (681)	G (990)	G (1000)	998	996	999
8	I (781)	G (1000)	G (986)	1000	1000	1000
9	I (263)	G (1000)	G (1000)	1000	1000	1000
10	I (883)	**I (610)**	**I (367)**	912	1000	996
11	I (887)	G (987)	G (1000)	1000	968	999
12	I (296)	G (958)	G (999)	999	999	937
13	I (796)	G (1000)	G (1000)	998	988	1000
14	**G (989)**	G (977)	G (987)	1000	1000	1000
15	I (3)	G (919)	**I (824)**	996	974	979
16	I (625)	G (1000)	G (976)	1000	1000	952
17		**I (889)**	G (958)		740	950
Mean	693	939	927	991	979	979
Accuracy	81.25%	82.35%	82.35%	100%	94%	100%
AUC	-	0.833	0.842	-	-	-

0.001, guilty: $t(16) = 7.054$, $p < 0.001$, CM: $t(16) = 5.218$, $p < 0.001$) (See Fig.4).

Bootstrap: Individual classification

Each participant was classified guilty (G) or innocent (IC) based on the bootstrapped amplitude difference (BAD) method, comparing the amplitude of P300 between probe and Iall stimuli. The cutoff for correct detection rate was 90%. The results are shown in Table 6. Three participants in the IC group were wrongly classified as guilty (81.25%); three participants in the G group were wrongly classified as innocent (82.35%); three participants in the CM group were wrongly classified as innocent (82.35%; bolded in Table 7).

In addition, the P300 amplitudes of target and Nall was compared to confirm the participants' full attention to the task. The target/Nall comparison was conducted not to judge guilty or innocent, but to confirm that target's P300 component is larger than nontarget's. The BAD method perfectly classified (100%) the participants in the IC and CM group. Only one participant in the G group was wrongly classified (94%).

ROC curve was calculated to evaluate the individual classification accuracies, using the bootstrap values of the innocent group as specificity, shown in Figure 6. The area under curve (AUC) for the G group was 0.833 (95% CI: 0.662 - 0.939) and for the CM group was 0.842 (95% CI: 0.673 - 0.945). There was no significant difference between the G and CM groups (p = 0.86).

Discussion

The present study was designed to revalidate the functionality of CTP in the setting with participant-oriented mental countermeasures and mock crime scenario. A behavioral measure of reaction times and a physiological measure of P300 components of ERP were analyzed. The results indicated that reaction times and P300 amplitudes can be a manifestation of countermeasure uses. Moreover, the CTP could correctly identify the IC, G, and CM groups with high accuracy. According results signify that the CTP can detect deception even when the countermeasures created based on the survey are used through analyzing P300 component. It has significance for including survey-based countermeasures since none of the related CTP studies have yet attempted them. The results can imply that the CTP not only works on specifically-assigned countermeasures, but also works on more flexible countermeasures where one can freely think of various countermeasures.

A reciprocal pattern of the reaction times between the G and CM group is consistent with the previous study (Rosenfeld et al., 2008). Participants in the guilty G group responded slower to the irrelevant probe compared to the irrelevant, whereas participants in the CM group responded faster to the probe than to the irrelevant. The result of the present study emphasized that the reaction times can be an indicator of the mental countermeasure use not only in autobiographical plot, but also in a mock crime scenario. However, the study conducted by Sokolovsky et al. (2011) posed a problem of reaction times as a marker of the countermeasure use. Participants were divided into simultaneous versus serial countermeasure users. When instructed to simultaneously use countermeasures and press the response button, the detection rate of reaction times greatly reduced. Moreover, an individual detection (guilty or innocent) using bootstrap method is necessary to precisely determine whether the reaction times can work as an indicator of the countermeasure use. The limitation of behavioral measures should not be ignored and more research examining the validity of reaction times should be conducted.

On the contrary to the reaction times, the P300 components has been discussed to be more effective in detecting deception and even be called the “brain fingerprints” (Farwell, Richardson, & Richardson, 2013). The results follow in the same manner: the P300 amplitudes worked as an indicator in discriminating the countermeasure uses. For over a decade, there has been an ongoing debate on whether a lie detector test can accurately detect lie or not. For example, the case law, such as *U.S. v. Urquidez* (1973) or *U.S. v. Scheffer* (1998), judged polygraph to be inadmissible, whereas the state of New Mexico passed the law allowing the submission of polygraph examination (NM Stat §29-14-5). The P300 components may provide a possibility for detection methods to be admissible in the court. Recent study related to the internal cost of deception using EEG spectral perturbations (Zhu et al., 2019) suggests future direction for the improvement of the field of the deception detection.

There have been mixed results of P300 for detecting the use of countermeasures. Some studies have distinguished countermeasure use by analyzing P300 component (Lukács et al., 2016), whereas some could not (Winograd & Rosenfeld, 2011). The present study might have been influenced by the “omit” effect, where a probe stimulus is the only stimulus that does not involve countermeasure and thus eliciting larger P300 component due to oddball effect (Meixner & Rosenfeld, 2010). Recently, a new ERP component, P900, has come up that it might act as an index of countermeasure use (Meixner et al., 2013). Component analysis of P900 with the present study’s data is necessary.

The survey was conducted prior to the main experiment to evaluate the opinions of respondents regarding the use of the countermeasures. Based on the post-experiment questionnaires investigating the used countermeasures, we could confirm that the countermeasures were properly performed by the participants. To our knowledge, a survey has not been widely used in the of deception. The results of the present study suggest the possibility of practical use of a survey in deception studies to convey participants’ opinion.

One of the limitations of the present study is that countermeasures were used for all irrelevant stimuli. Countermeasures were used for all irrelevant stimuli as the survey result indicated that they would widely use countermeasures for irrelevant stimuli. However, many studies have partially-assigned countermeasures for irrelevant stimuli, generally two out of four irrelevant stimuli, to increase the accuracy of the result (Hu, Hegeman, Landry, & Rosenfeld, 2012; Meixner et al, 2009; Meixner & Rosenfeld, 2010). Therefore, a future research should compare the difference of P300 component between all-assigned countermeasures and partially-assigned countermeasures with same setting of the present study.

Another limitation is that the exact countermeasures of Rosenfeld et al.'s (2008) study was not used in the current study and that a mock crime was performed in laboratory settings. Although mock crime can have high ecological validity, it has constraints in aspects of controllability and reality. Further research should be directed toward creating a setting close to an actual crime scene by using tools such as virtual reality.

Overall, the results of the present study corroborate previous results by Rosenfeld and his colleagues and could successfully replicate the CTP with high detection accuracy. In particular, it should be highlighted that the present study is the first, as far as we know, to directly reflect survey data to create a survey-based, participant-oriented countermeasures for the CTP and to conduct a mock crime prior to the CTP where all mental countermeasures were used. Further studies using source and connectivity analyses should be considered to elaborate the underlying neural mechanism of countermeasures.

References

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*, 504-522.

Beck, A. T., Emery, G., & Greenberg, R. L. (2005). *Anxiety disorders and phobias: A cognitive perspective* (15th ed.). New York: Basic Books.

Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: An overview. *Frontiers in Psychology*, *3*, 1-11.

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*(3), 273-281.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1982). An evaluation of polygraphers' judgments: A review from a decision theoretic perspective. *Journal of Applied Psychology*, *67*(6), 701-713.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*(2), 319-333.

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47-63.

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. (L. Festinger & S. Schachter, Eds.). Academic Press.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357-374.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456-467.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, *28*(6), 632-638.

Farwell, L. A., Richardson, D. C., & Richardson, G. M. (2013). Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information. *Cognitive Neurodynamics*, 7(4), 263-299.

Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67(2), 131-137.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Laufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2. Cognitive processes and performance* (pp. 1-49). New York: John Wiley & Sons.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, *79*(2), 252-259.

Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, *34*(1), 108-115.

Hu, X., Hegeman, D., Landry, E., & Rosenfeld, J. P. (2012). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology*, *49*(1), 85-95.

Hu, X., & Rosenfeld, J. P. (2012). Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. *Psychophysiology*, *49*(8), 1090-1100.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4-5), 411-430.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*(3), 259-273.

Knowles, E. S., Cutter, H. S. G., Walsh, D. H., & Casey, N. A. (1973). Risk-taking as a personality trait. *Social Behavior and Personality: An International Journal*, *1*(2), 123-136.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and*

Performance, *11*(4), 393-408.

Kramer, A., Wickens, C., Vanasse, L., Heffley, E., & Donchin, E. (1981). Primary and secondary task analysis of step tracking: An event-related potentials approach. *Proceedings of the Human Factors Society Annual Meeting*, *25*(1), 43-47.

Loeb, A., Beck, A. T., Diggory, j. C., & Tuthill, R. (1967). Expectancy, level of aspiration, performance, and self-evaluation in depression. In *Proceedings of the 75th Annual Convention of the American Psychological Association* (Vol. 2, pp. 193-194).

Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56-65.

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.

Lykken, D. T. (1984). Detecting Deception in 1984. *American Behavioral Scientist*, *27*(4), 481-499.

Marino, C. F., Reid, J. E., & Inbau, F. E. (1978). Truth and deception: The polygraph ("lie-detector") technique. *The Journal of Criminal Law and Criminology*.

Marston, W. M. (1917). Systolic blood pressure symptoms of deception. *Journal of Experimental Psychology*, *2*(2), 117-163.

Meixner, J. B., Haynes, A., Winograd, M. R., Brown, J., & Rosenfeld, J. P. (2009). Assigned versus random, countermeasure-like responses in the P300 based complex trial protocol for detection of deception: Task demand effects. *Applied Psychophysiology Biofeedback*, *34*(3), 209-220.

Meixner, J. B., Labkovsky, E., Peter Rosenfeld, J., Winograd, M., Sokolovsky, A., Weishaar, J., & Ullmann, T. (2013). P900: A putative novel ERP component that indexes countermeasure use in the P300-based concealed information test. *Applied Psychophysiology Biofeedback*, *38*(2), 121-132.

Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based concealed information test. *Psychophysiology*, *47*(1), 57-65.

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*(2), 286-298.

National Research Council, Committee to Review the Scientific Evidence on the Polygraph (2003). *The polygraph and lie detection*. Washington, DC: The National Academies Press.

Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2012). AMICA: An adaptive mixture of independent component analyzers with shared components. *Swartz Center for Computational Neuroscience*, (December 2014), 1-15.

Park, H. D., Correia, S., Ducorps, A., &

Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience*, *17*, 612-618.

Patrick, C. J., & Iacono, W. G. (1991). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology*, *28*(6), 632-638.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin*, *84*(4), 782-799.

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

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 247-296). Springer Publishing Co.

Reid, J. E. (1945). Simulated blood pressure responses in lie-detector tests and a method for their detection. *Journal of Criminal Law* & *Criminology*, *36*(3), 201-214.

Rosenfeld, J. P. (2002). Event-related potentials in detection of deception, malingering, and false memories. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 265-286). New York: Academic 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), 251-259.

Rosenfeld, J. P., & Labkovsky, E. (2010). New P300-based protocol to detect concealed information: Resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. *Psychophysiology*, *47*(6), 1002-1010.

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, *45*(6), 906-919.

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*(2), 205-219.

Rosenfeld, J. P., Tang, M., Meixner, J., Winograd, M., & Labkovsky, E. (2009). The effects of asymmetric vs. symmetric probability of targets following probe and irrelevant stimuli in the complex trial protocol for detection of concealed information with P300. *Physiology and Behavior*, *98*(1-2), 10-16.

Rosenfeld, J. P., Ward, A., Thai, M., & Labkovsky, E. (2015). Superiority of pictorial versus verbal presentation and initial exposure in the P300-based, Complex Trial Protocol for concealed memory detection. *Applied Psychophysiology Biofeedback*, *40*(2), 61-73.

Sasaki, M., Hira, S., & Matsuda, T. (2001). Effects of a mental countermeasure on the physiological detection of deception using the event-related brain potentials. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *72*(4), 322-328.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist*, *40*(3), 355-366.

Schoonjans, F., Zalata, A., Depuydt, C. E., & Comhaire, F. H. (1995). MedCalc: A new computer program for medical statistics. *Computer Methods and Programs in Biomedicine*, *48*(3), 257-262.

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, *51*(1), 125-139.

Sokolovsky, A., Rothenberg, J., Labkovsky, E., Meixner, J., & Rosenfeld, J. P. (2011). A novel countermeasure against the reaction time index of countermeasure use in the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *81*(1), 60-63.

Song, I., Kim, H., Lee, K. E., Chang, E., & Kim, H. T. (2019). Can virtual mock crime replace actual mock crime? An event-related potential study. *The Korean Journal of Forensic Psychology*, *10*(2), 89-109.

Soskins, M., Rosenfeld, J., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: Complex vs. simple paradigms. *International Journal of Psychophysiology*, *40*(2), 173-180.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, *155*(3768), 1436-1439.

United States *v.* Scheffer, 523 U.S. 303 (1998)

United States *v.* Urquidez, 356 F. Supp. 1363 (C.D. Cal. 1973)

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping applications to psychophysiology. *Psychophysiology*, *26*(2), 208-221.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (pp. 4101-4105).

Winograd, M. R., & Rosenfeld, J. P. (2011). Mock crime application of the Complex Trial Protocol (CTP) P300-based concealed information test. *Psychophysiology*, *48*(2), 155-161.

Zhu, C., Pan, J., Li, S., Liu, X., Wang, P., & Li, J. (2019). Internal cost of spontaneous deception revealed by ERPs and EEG spectral perturbations. *Scientific Reports*, *9*(1), 1-11.

1 차원고접수 : 2019. 12. 27.
심사통과접수 : 2020. 03. 24.
최종원고접수 : 2020. 03. 26.

한국심리학회지 : 법
The Korean Journal of Forensic Psychology
2020, Vol. 11, No. 1, 89-115

설문 기반 대응방안을 사용한 복합시행 프로토콜의 재평가

김 혜 민 송 인 욱 장 은 희 김 현 택

고려대학교 심리학과

전통적인 거짓말 탐지 방법은 범죄와 관련 있는 질문과 관련 없는 질문을 할 때 범죄자의 자율 신경 반응을 비교하여 유무죄를 판정한다. 거짓말 탐지 검사의 정확성과 유효성을 높이기 위해 연구자들은 효과적인 대안을 꾸준히 모색하였다. 그중 P300 기반 Complex Trial Protocol(CTP)은 탐침 자극과 무관련 자극의 P300 진폭을 비교하여 유무죄를 판정하는 방법으로, 이전 거짓말 탐지 방법의 취약점인 대응방안을 사용하는 경우에도 유무죄 판별의 정확도가 유지된다는 장점이 있다. 하지만 대부분의 기존 연구는 Rosenfeld et al.(2008)이 실험에 적용한 대응방안을 사용했다는 문제점이 있다. 본 연구는 먼저 대응방안에 대한 설문 조사를 실행하여 이에 관한 결과를 본 실험의 대응방안으로 사용하였다. 그리고 이러한 대응방안에 대해서도 CTP 기법이 범죄 사실 여부를 정확히 판별할 수 있는지 확인하고자 하였다. 총 50명의 참가자를 세 집단(무죄 집단, 유죄 집단, 대응방안 집단)으로 나누고 CTP를 수행하게 하였다. 대응방안 집단은 CTP를 수행하는 동안 검사자 모르게 앞서 선정된 대응방안을 사용하였다. P300 진폭 분석 결과, 유죄 집단의 탐침 자극 진폭이 무관련 자극 진폭보다 유의미하게 큰 것으로 나타났다. 대응방안 집단 역시 대응방안을 사용했음에도 불구하고 탐침 자극에서 유의미하게 큰 P300 진폭을 보였다. 진폭 부트스트랩 분석(BAD) 결과 유죄 집단, 무죄 집단, 대응방안 집단의 유무죄 판정 정확도는 각각 81.25%, 82.35%, 82.35%였다. 이러한 결과는 CTP가 설문 조사를 기반으로 한 대응방안에 대해서도 높은 범죄 탐지율을 보이며 향후 실제 수사 현장에서 사용될 수 있는 가능성을 시사한다.

주요어 : 거짓말탐지, 복합시행 프로토콜, P300, 대응방안