

A CONCEALED INFORMATION TEST WITH COMBINATION OF ERP RECORDING AND AUTONOMIC MEASUREMENTS

Received October 31, 2012.

The concealed information test (CIT) is based on a comparison between the subject's physiological responses to a probe and irrelevant items in order to detect concealed information. The main purpose of our study was to investigate the CIT accuracy enhancement related to a combination of recording of event-related potentials (ERPs) and autonomic measurements. We tried to maximally liken the experimental conditions to real ones by the use of a criminal context in the "mock crime" instruction and real innocent subjects instead of hypothetical ones. Fifty-two subjects volunteered and performed just one of the innocent or guilty scenarios. The CIT was designed in five blocks with short interstimulus intervals. In each block, stimuli were presented in the 7th-order balanced sequence. In addition to EEG phenomena, the heart rate, skin conductance responses (SCRs), respiratory activity, and finger plethysmogram were recorded. Statistical analyses showed that there was a significant difference between standardized difference scores of the guilty and innocent groups in both ERP and autonomic measures. The SCR did not achieve the expected results reported in standard autonomic-based CIT studies. A review of the two classification methods showed that the combination of ERP and autonomic measurements enhances the CIT accuracy. The best classification accuracy obtained by the aid of linear discriminant analysis (LDA) was 90.9%. It seems that using a criminal context in the "mock crime" instruction and the reward-punishment system made subjects more attentive and involved in the experiment; therefore, the accuracy was improved compared with that in similar studies.

Keywords: concealed information test, event-related potentials, autonomic responses, logistic regression model, linear discriminant analysis.

INTRODUCTION

Attempts to detect concealed information using recording of physiological indices have a rather long history. Initial studies in this matter refer to understanding the relationship between the heart rate (HR) and deceptively denying knowledge. Further studies (such as Lombroso in the late 19th century and Marston in 1917) were carried out to find new deception signs [1-3]. The most common physiological measures in polygraph systems are parameters of respiration, cardiovascular measures, and electrodermal responses, which mainly reflect functions of the autonomic nervous system (ANS).

In recent years, various approaches were introduced for psychophysiological detection of deception, such as studying brain functions in a deception procedure using functional brain imaging and also recording and investigation of brain potentials [4, 5]. Event-related brain potentials (ERPs) were widely studied and demonstrated more satisfactory results [6]. The P300 wave is the most important component of ERPs, which is recently used in most studies. In some previous ERP studies, the P300 amplitude was reported to be a reliable index for detection of deception [7, 8]. In recent studies, a new approach was introduced in order to improve the results of detection of deception by combining the measurements of ANS and CNS functioning.

The concealed information test (CIT), also referred to as the guilty knowledge test [9], is an effective method of psychophysiological detection of concealed information on crime [10]. In this method, differential

¹ Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran.

Correspondence should be addressed to M. H. Moradi (e-mail: mhmoradi@aut.ac.ir).

physiological responses to specific items are surveyed [9]. One of these items (probe) corresponds to the aspects of the crime that are under investigation, and the other one is an irrelevant item. The irrelevant and probe items are repeatedly presented in a certain sequence. An innocent subject without any knowledge of the crime demonstrates similar physiological responses to both items, whereas a guilty subject who deceptively denies his deed-related knowledge shows different physiological responses to these items [9].

A CNS response that mirrors cognitive processing and a peripheral response that mainly reflects a function of the ANS might complement each other more effectively. However, a combination of ERP recording and autonomic measures within the same experiment generally entails some difficulties [11]. First, short interstimulus intervals (ISIs) should be used in ERP-based tests because it was shown that the P300 amplitude is affected by ISI values [12]. At the same time, long ISIs (20-30 sec) are used in autonomic-based CITs to provide an adequate recovery time. Measuring the skin conductance response (SCR) within a short-ISI paradigm results in overlapping responses that are difficult to quantify independently [11, 13]. Second, EEG evaluation based on a single trial is not very reliable due to a rather low signal-to-noise ratio. Therefore, large numbers of stimuli are presented in most ERP studies in order to obtain an adequate number of valid ERPs per condition [11]. In contrast, the autonomic-based CIT uses smaller numbers of stimulus presentations because the amplitude of autonomic responses becomes critically small [13], while the SCR is known to intensely habituate when a large number of stimuli are presented [11].

However, a few recent studies have combined autonomic and ERP measurements in the detection of concealed information [11]. Studies in this field are divided into two categories. The first category includes studies where short ISIs and large numbers of stimulus presentations were used [11, 14]. The second category includes studies with long ISIs and limited numbers of stimulus presentations [15, 16]. In the second category, the studies, except that by Matsuda et al. [17], did not achieve a typical probe vs irrelevant difference for the P300 amplitude. The authors believed that using long ISIs was the reason for this phenomenon and suggested that further studies should evaluate shorter ISIs as a solution [16]. In the first category, studies gained incremental validity from combined measurements. Due to using short

ISIs and large numbers of stimulus presentations, the discrimination and correct classification rate of autonomic measures in these studies remained below those in most autonomic-based CIT studies.

The purpose of our study was to evaluate the CIT accuracy enhancement with the aid of combination of ERP recording and autonomic measures. In this study, short ISIs and large numbers of stimulus presentations are primarily selected as the paradigm. On the other hand, some arrangements should be considered to handle the side effect of short ISIs in autonomic measurements. The second purpose of the study was to make the experimental conditions closer to real conditions. This purpose has been achieved in two ways, first by designing a “mock crime” scenario with a criminal context, and, second, by designing an innocent scenario and involving real innocent subjects instead of hypothetical ones.

METHODS

Participants. Fifty-two healthy students of the Biomedical Engineering Faculty in the Amirkabir University of Technology (40 men and 12 women; mean age 22.5 ± 3.5 years; all right-handed; all had normal or corrected vision) participated in this experiment. They were rewarded with a gold coin (value around US \$12) after the experiment.

Design and Procedure. After the subjects had given their informed consent, they were asked to select one of two envelopes containing the instruction of a “guilty or innocent” scenario. It should be taken into account that both of envelopes contained the same instruction, but this instruction was referred to the guilty scenario for half of the subjects, while the scenario was innocent for another half.

After selection of the scenario, the experimenter left the laboratory, and the subjects read the instruction in order to perform the respective actions. The guilty scenario consisted of stealing a gold coin (value around US \$12, hidden in a wallet) and a cell phone from a personal locker in the laboratory. In order to open the personal locker, subjects had to find a key hidden in a cupboard in the kitchen. In the innocent scenario, the task involved washing dirty cups put in the kitchen sink. At the end of both scenarios, subjects were requested to go to the lobby and wait for the experimenter to come. After 7 min, the experimenter approached the subjects and informed them that “a crime has been committed, and you are one of the

suspects; if you succeed in taking the test, you will obtain a gold coin.” Following this, the experimenter escorted the subjects back to the laboratory, where the main test was performed. Because the subjects did not know that the two envelopes contained the same instruction, and they were trying to get the gold coin, it was expected that the subjects were under stress. These arrangements likened the mock crime scenario to real condition and prevented the subjects from being inattentive.

In this study, a variant of the CIT consisting the five blocks was used. The presented pictures in each block belong to one category of the details of the crime scene (i.e., coins, keys, cell phones, wallets, and lockers). Each block contained seven types of the stimuli (one target, one probe, four irrelevant, and a null event). Each stimulus was presented seven times except for the null stimulus that was presented six times. The sequence of the presented stimuli was based on a pseudorandom sequence called M-sequence that is a balanced-order sequence [18]. Fourteen stimuli were added to the sequence of 48 ones, creating a “history” and “future” for the initial and last stimuli. Thus, the stimuli were presented in a pseudorandom series of 62 stimuli in each block. These stimuli were presented with ISIs varying from 2.3 to 2.7 sec. One-minute-long rest between blocks were set. The designed CIT is illustrated in Fig. 1. A 17-color screen at a distance of 90 cm from the subject was used for presentation of

the stimuli.

To overcome the overlap problem (in particular, overlapping SCRs) in our study, we used the 7th-order M-sequence for each block of the CIT. For this purpose, a null stimulus was added to the standard stimulus set (Fig. 1). The characteristic of the M-sequences is that each type of the stimulus has an identical bias. This means that, although the average response for each type is biased by responses to previous stimuli, this bias is identical for every stimulus type [14]. So, the responses can be compared without worrying about an overlap problem. As an example of the 3rd-order M-sequence, see Fig. 2. This sequence has three types of stimuli. Each stimulus was repeated three times, except for C (a null stimulus was added to stimulus set) that was repeated two times. Figures 2b and 2c show the bias on the A and B, respectively. Due to the characteristic of M-sequence, both A and B have an identical bias.

The subjects were randomly divided into guilty and innocent groups. At least 26 subjects were assigned to the guilty group and performed the guilty scenario. The other subjects were assigned to the innocent group and performed the corresponding scenario. The items presented to the subjects of both groups were identical. A few test results were removed due to misdoing of the protocol or inappropriately recorded signals. Finally, 23 guilty subjects and 21 innocent subjects were chosen for use the subsequent data analyses.

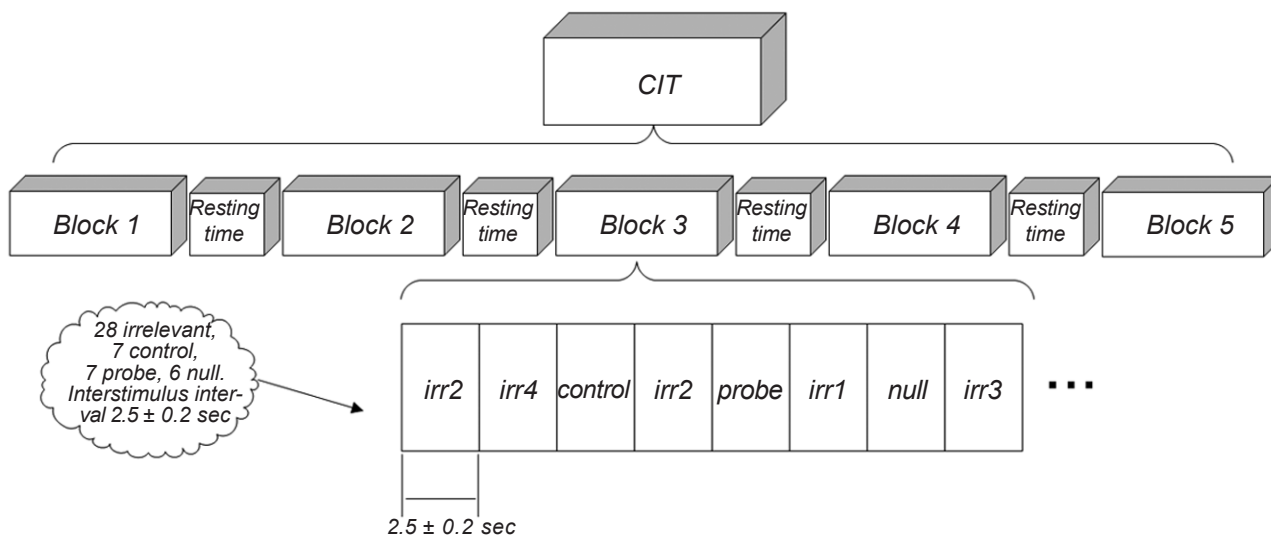


Fig. 1. The designed CIT. Control, probe, irr1, irr2, irr3, and irr4 are control stimulus, probe stimulus, and four irrelevant stimuli.

Р и с. 1. Схема організації тесту з прихованою інформацією.

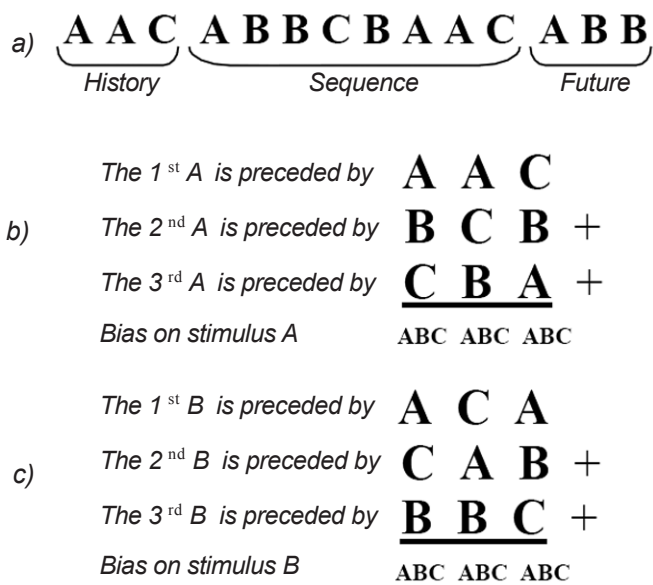


Fig. 2. The third-order M-sequence. a) Sequence of the stimuli, b) bias on stimulus “A,” and c) bias on stimulus “B” are illustrated.

Рис. 2. М-последовательность третьего порядка.

In the first step of the test, EEG electrodes, electrooculogram (EOG) electrodes, and leads for the peripheral measurements were attached, and the subjects were instructed to attend to all presented items and acknowledge recognition of the target picture by right clicking the mouse, while left clicking for all pictures meant that they were unable to recognize.

Physiological Recording. The procedure of physiological recording took place in a silent environment (laboratory) and in the absence of other people.

The EEG data were recorded using eight active Ag/AgCl electrodes with the gUSBamp system (G.Tec, Austria). Electrodes were placed at Fz, Cz, Pz, Oz, C4, C3, P4, and P3 sites according to the international 10-20 system and referenced to an electrode at the left earlobe. For controlling eye movements, vertical and horizontal EOGs were recorded. The EEG and EOG data were digitized at 256 sec⁻¹ and filtered online using a 0.1-30 Hz bandpass and a 50-Hz notch filter.

The skin conductance changes were recorded by two electrodes via an isolated amplifier (MLT 116F and FE116, respectively; ADInstrument, Australia) with low-voltage 75-Hz alternating current. Electrodes were placed on the volar side of the middle phalanges of the index and fourth fingers of the left hand.

The finger plethysmogram signal was recorded using

an infrared system in a spring clip (MLT1020F) via an isolated amplifier (ML110; both by ADInstrument, Australia) from the middle fingertip of the left hand.

The thoracic and abdominal respiratory activities were recorded using two piezo respiratory belt transducers (MLT1132, ADInstrument, Australia) generating a voltage when there is a change in the thoracic or abdominal circumference due to respiration. All peripheral signals were digitized with a sampling rate of 10³ sec⁻¹.

Data Analysis. After filtering the signals, we separated each continuous record into single sweeps according to the known onset times of the stimulus presentation. The EOG data were checked for blink artifacts by visual inspection, and sweeps with such artifacts were removed. The ERPs for each type of the stimuli (probe, or target, or irrelevant) were separately extracted by averaging between related single sweeps. The P300 peak-to-peak amplitude was measured. In this measurement, a maximally positive segment average of 100 msec was searched within a window from 400 to 900 msec after the stimulus. The midpoint of the maximum positivity defined the P300 latency. After that, the algorithm searched for the maximum 100-msec-long negativity within the window from the P300 latency to the end of the sweep. The difference between the maximum positivity and negativity defined the peak-to-peak measure [19].

The SCR is one of the slow responses of the ANS. It has an onset latency and a rise time of 1 to 3 sec and a half-recovery time of up to 10 sec [20]. When stimuli are presented with short intervals, the current response is influenced by previous responses. In other words, the responses overlap. To overcome the problem of overlapping in SCRs in our study, the stimuli were presented in a 7th-order M-sequence, as was mentioned above. The SCR was assessed based on the averaging method proposed by Meijer [14]. The epochs were extracted from -1 to 20 sec relative to the stimulus onset and baseline corrected at the sample preceding the stimulus onset. Within each block, these epochs were averaged per stimulus type. Since no picture was presented at presentation of the null stimulus, an estimate of the bias produced by the response to the previous stimuli was obtained by averaging on this event [14]. Thus, by subtracting it from other responses, we can assume that they are unbiased. So, the average of the null stimulus was subtracted from the average of each stimulus type in each block.

The respiration line length (RLL) is a useful

measurement for detection of deception that integrates information on the frequency and depth of respiration. In our study, the RLL was automatically computed over a time window from 600 to 3000 msec post-stimulus onset; the method was derived from Timm [21, 22]. The data from abdominal and thoracic channels were averaged.

The phasic heart rate (pHR) was calculated based on the HR. The HR was defined based on the *R-R* intervals of the ECG signal. The AC component of the photoplethysmogram (PPG) pulse is synchronous with the heart beat and, therefore, can also be a source of HR information [23]. In our study, PPG peaks were automatically detected based on an adaptive threshold method [24]. Peak-to-peak intervals were transformed into the HR and real-time scaled [25]. The HR during the last second before the stimulus onset served as a prestimulus baseline. The pHR values were defined by subtracting this baseline value from each second-per-second poststimulus value [11]. For extracting the trial-wise information of pHR, the mean change in the HR within 3 sec after the stimulus onset, compared with the prestimulus baseline, was calculated [26].

In order to eliminate individual differences in the responsiveness, physiological and behavioral measures should be standardized [27]. Z-transformed values were calculated for each subject and each block. All probe and irrelevant trials of one block (not including 14 trials of “history” and “future” in each block) were used for calculation of individual means and standard deviations [14]. In the detection of the deception procedure, the difference between responses to probe and irrelevant items is a basic indicator. Thus, difference scores were calculated as proposed by Gamer et al. [28]. In this method, the difference score is defined as the difference between the mean of the standardized probe trials and the mean of all standardized irrelevant trials within each block. Afterwards, the mean of five blocks was computed as an overall index of the differential responsiveness in each physiological or behavioral measure. These values were used in subsequent statistical analyses.

Statistical Analysis. The cross-correlation coefficient is a reliable feature that is studied in the detection of deception by means of a bootstrapped correlation difference (BCD) method. The BCD answers the question: “Are the cross-correlation coefficients between ERP responses to probe and target stimuli significantly greater than the corresponding cross-correlation of responses to probe and irrelevant stimuli?” If so, the subject is found to be guilty [5].

The statistical technique of bootstrapping [29] shows the statistical significance of this hypothesis. In our study, the BCD method was applied to the artifact-free single sweeps as proposed by Abootalebi et al. [5]. The output parameter of the BCD method (N_{D-0}) means that the probe response is more different from the irrelevant and more similar to the target response; thus, this subject is more likely to be guilty, and *vice versa*. The N_{D-0} value determined for each subject was used in the subsequent statistical analyses.

The standardized difference scores of autonomic and ERP responses were compared between the guilty and innocent groups using the ANOVA test. The significance level for the assessment of main and interaction effects was set to 0.05. The Cohen’s *d* was calculated as an estimate of the effect size [30, 31].

Classification. To achieve applicable aspects of this study, it is necessary that the subjects be classified into two groups, innocent and guilty. In order to find an optimized combination of the ERP and autonomic measures, the discrimination performance of each measure and combination of measures were evaluated using two methods, a binomial logistic regression model and linear discriminant analysis.

The logistic regression model is used extensively in medical and social science fields as a basic method for describing the relationship between a response variable and one or more explanatory variables. The goal of an analysis using the logistic regression model is to find the best fitting and most parsimonious, yet biologically reasonable, model to describe the relationship between an outcome (dependent, or response) variable and a set of independent (predictor, or explanatory) variables [32].

Linear discriminant analysis (LDA) is a commonly used technique for data classification and dimensionality reduction. This method maximizes the ratio of the between-class variance to the within-class variance in any particular data set, thereby guaranteeing maximal separability [33, 34]. The aim of LDA (also known as Fisher’s LDA) is to use hyperplanes to separate the data representing different classes. For a two-class problem, the class of a feature vector depends on which side of the hyperplane the vector is [35]. This technique is characterized by very low computational requirements and generally provides good results, which makes it suitable for many pattern recognition problems.

The performance of the logistic regression model and LDA was estimated using the leave-one-out cross-validation method. Each subject, once excluded from

TABLE 1. Means and Standard Deviations (s.d.) of the Standardized Difference Score of Autonomic and ERP Measures in the Guilty and Innocent Groups

Т а б л и ц я 1. Середні значення та стандартні відхилення стандартизованих різниць оцінок вегетативних показників та параметрів пов'язаних з подією потенціалів у групах „винуватих” та „невинуватих” суб'єктів

Measures	Guilty group		Innocent group	
	means	s.d.	means	s.d.
P300	0.60	1.20	-0.05	0.81
BCD	75.65	22.91	23.38	16.41
pHR	-0.18	0.24	0.08	0.21
SCR	0.43	0.86	-0.24	0.84
RLL	-0.13	0.41	0.08	0.15

Footnote. P300 is the most important component of ERP; BCD is the bootstrapped correlation difference; pHR is the phasic heart rate; SCR is the skin conductance response, and RLL is the respiration line length.

the original data, was set as test data, and the other subjects were used as training data. In each iteration, the classifier was trained for using training subject data and their real labels. After that, the held-out subject was classified as guilty or innocent with the trained classifier. Finally, the accuracy was calculated based on the classification result of the held-out subjects.

RESULTS

The means and s.d. of standardized difference scores

of autonomic and ERP responses in the guilty and innocent groups are summarized in Table 1.

Event-Related Potentials. All statistical analyses were performed on the Pz channel, where the P300 amplitude is typically the largest. Figure 3 shows grand means of the ERP waveforms for probe, target, and irrelevant stimuli in the guilty and innocent groups (A and B, respectively) for 1,000 msec after the stimulus. As was expected, a large positivity was elicited by the target stimuli but not by the irrelevant stimuli. In the guilty group (A), probe responses demonstrated similarity with target ones, while some

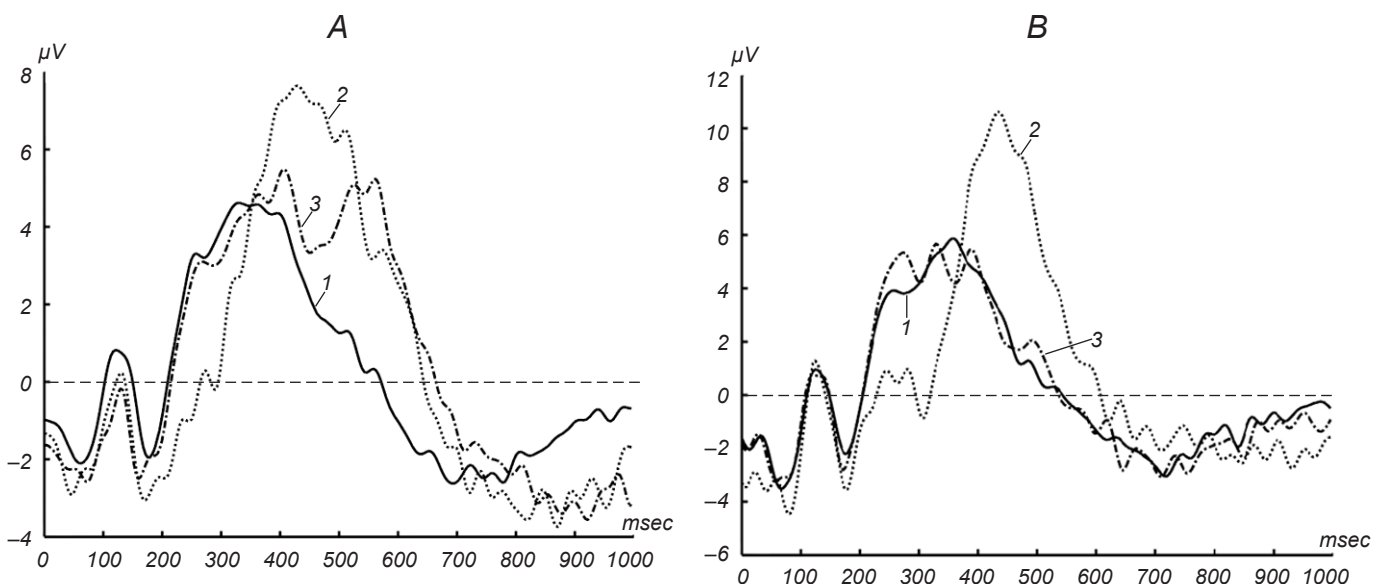


Fig. 3. Grand means of the ERP waveforms for irrelevant stimuli (1), target stimuli (2) and probe stimuli (3) in the guilty (A) and innocent (B) group. Abscissa) Time, msec; ordinate) amplitude, μV .

Р и с. 3. Усредненні пов'язані з подією потенціали при пред'явленні іррелевантних (1), цільових (2) та зондуючих (3) стимулів у групах „винуватих” (A) та „невинуватих” (B) тестованих суб'єктів.

striking similarities between the probe and irrelevant responses were observed in the innocent group (B).

The standardization procedure of P300 amplitude was so different from the other measures. As was explained, the ERP was extracted from all single sweeps of each type of stimulus. Thus, since the within-block standardization was impossible, the within-subject standardization was calculated based on the mean and s.d. of the P300 amplitude of probe and irrelevant ERPs. ANOVA for the P300 amplitude data showed that this parameter observed in the guilty group was significantly greater than that in the innocent group ($F = 4.32$; $P < 0.043$; $d = 0.62$).

As was explained, the percentage of the probe responses, which was more similar to target than to irrelevant ones, was computed as the BCD. In the BCD, before computing the cross-correlation coefficients, a time window was applied to single sweeps between 300 and 900 msec after stimulus, and correlation of the sweeps was only noticed in this time-limited interval, because we expected that the P300 would appear exclusively in this region [5]. Only the correlation coefficient at lag = 0 was considered. Since there is no individual difference in the BCD measure, there is no need for standardization. The ANOVA test between two groups showed that the BCD measure was significantly greater in the guilty group ($F = 74.39$, $P < 0.001$, $d = 2.60$). Since the difference between the guilty and innocent groups in the BCD was much more significant than that in the P300 amplitude, only the BCD was evaluated as the ERP measure in the subsequent analyses.

Autonomic Responses. To examine the statistical

distribution of autonomic responses in the guilty and innocent groups, the box plot of the standardized difference score of these responses was used (Fig. 4.).

ANOVA for the pHR data showed that these values in the innocent group were significantly greater than those in the guilty group ($F = 15.04$; $P < 0.001$; $d = -1.17$; Fig. 4A).

ANOVA for the RLL data showed that these values in the innocent group were significantly greater than those in the guilty group ($F = 5.13$; $P = 0.028$; $d = -0.68$; Fig. 4B). As can be seen in this figure, the box plots are very compact, and there is good discrimination between the guilty and innocent groups.

Assessment of the SCR using ANOVA showed that the respective values in the guilty group were significantly greater than those in the innocent group ($F = 6.95$; $P = 0.011$; $d = 0.79$; Fig. 4C).

Logistic Regression Model. To compare the discrimination performance of the ERP and the autonomic measures, different combinations of these measures were evaluated. Subjects were classified as guilty or innocent based on a criterion $P > 0.5$ for classification as guilty. Figure 5 shows the correct-classification rates based on the leave-one-out cross-validation method using BCD, SCR, pHR, RLL, and other combinations of measures as predictors.

Evaluation of different independent measures shows that BCD gives the best performance with 88.63% correct-classification rate, and RLL with 79.54% is the best autonomic measure.

As can be seen in Fig. 5 and Table 2, the best correct-classification rate of different combinations of measures is 88.63%, which was obtained in the



Fig. 4. Box plot of the standardized difference scores of autonomic measures in the guilty and innocent groups. A) Phasic heart rate (pHR), B) respiration line length (RLL), and C) skin conductance response (SCR).

Рис. 4. Бокс-діаграми стандартизованих різниць бальних оцінок вегетативних показників у групах „винуватих” та „невинуватих” суб’єктів.

TABLE 2. Best Correct-Classification Rates (Accuracy) of a Logistic Regression Model and Linear Discriminant Analysis Using Event-Related Potential (ERP) Measures, Autonomic Measures, and Combination of the Measures

Т а б л и ц я 2. Найкращі рівні коректної класифікації (точності) для моделі логістичної регресії та лінійного дискримінантного аналізу з використанням вимірів параметрів пов'язаних з подією потенціалів, вегетативних показників та комбінації таких показників

Methods	ERP measure, %	Autonomic measures, %	Combination of measures, %
Logistic regression model	88.63	79.54	88.63
Linear discriminant analysis	88.63	79.54	90.9

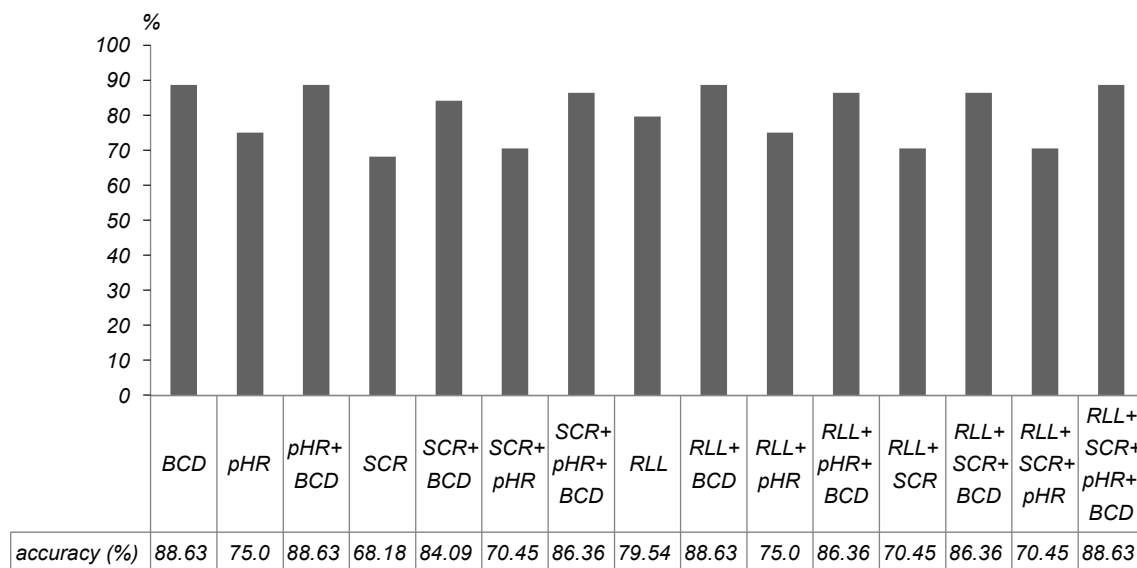


Fig. 5. Correct-classification rates (accuracy) for classification of the subjects as guilty or innocent using a logistic regression model based on the leave-one-out cross-validation method. Bootstrapped correlation difference (BCD), phasic heart rate (pHR), skin conductance response (SCR), respiration line length (RLL), and other combinations of the measures are used as predictors.

Р и с. 5. Рівні коректної класифікації (точності) при поділі тестованих суб'єктів на „винуватих” та „невинуватих” з використанням моделі логістичної регресії.

tree combinations (pHR + BCD, RLL + BCD, and RLL + SCR + pHR + BCD). None of the combination of the ERP and the autonomic measure yielded the incremental validity.

Linear Discriminant Analysis (LDA). Figure 6 shows the correct-classification rate of LDA based on the leave-one-out cross-validation method using BCD, SCR, pHR, RLL, and other combinations of the measures as features. As can be seen in Fig. 6 and Table 2, evaluation of different independent measures shows that the BCD gives the best performance with an 88.63% correct-classification rate, and the RLL at 79.54% is the best autonomic measure.

Evaluation of different combinations of the measures

using LDA shows that the best correct-classification rate is 90.9% (Table 2). According to Fig. 6, the incremental validity can be seen in two cases with a 90.9% correct-classification rate (pHR + SCR + BCD and pHR + SCR + RLL + BCD).

DISCUSSION

In this study, a variant of the CIT based on the simultaneous use of EEG and autonomic measurements was designed. Recent publications emphasized the need for studying the combination of EEG and autonomic measurements in the detection of

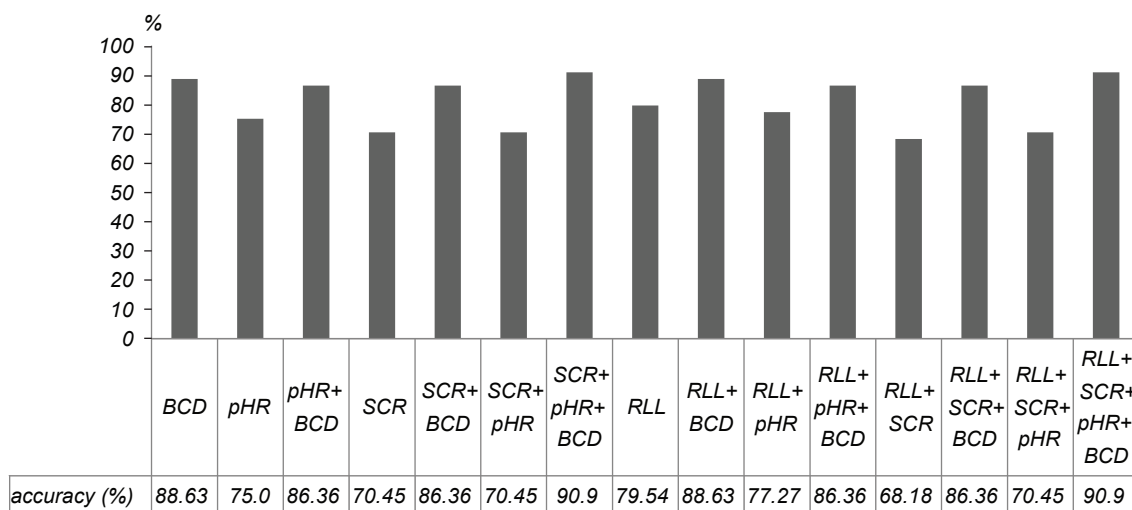


Fig. 6. Correct-classification rates (accuracy) for classification of the subjects as guilty or innocent using linear discriminant analysis based on the leave-one-out cross-validation method. Bootstrapped correlation difference (BCD), phasic heart rate (pHR), skin conductance response (SCR), respiration line length (RLL), and other combinations of the measures are used as features.

Рис. 6. Рівні коректної класифікації (точності) при поділі тестованих суб'єктів на „винуватих” та „невинуватих” з використанням лінійного дискримінантного аналізу.

deception [11, 15-17]. The purpose of our study was to evaluate the CIT accuracy enhancement caused by the combination of ERP and autonomic measurements. The second purpose was likening of the experimental conditions to real conditions. So, the CIT was tested on subjects who performed under the innocent and guilty scenarios.

Statistical analyses (summarized in Table 1) showed that both in the ERP and autonomic measures, there is a significant difference between the standardized difference scores in the guilty and innocent groups. Due to the use of short ISIs, this significant difference was predictable for the ERP measure, while the significant difference for the autonomic measures means that the considered arrangements allowing us to solve the problems of habituation and overlapping have been successful. These arrangements included the five-block CIT, variable ISIs, and stimulus presentation in the balanced order (using M-sequence series). The five-block CIT and variable ISIs were utilized to solve the problem of habituation; stimulus presentation in the balanced order was used to overcome overlapping in the SCRs.

The difference between innocent and guilty groups in the BCD is significantly greater than that in the P300 amplitude. Figure 3 shows that there are some striking similarities between the probe and irrelevant responses in the innocent group in contrast to the guilty group.

This result is consistent with our expectation and also with the previous ERP-based CIT studies [6, 8, 36].

Statistical analyses of autonomic measures show that there are significant differences between standardized difference score in the guilty and innocent groups. More detailed examination shows that the pHR has the best efficiency among autonomic measures, while most autonomic-based studies have introduced the SCR as the best measurement for detection of deception. The lower significance of the SCR might be due to overlapping caused by short ISIs. Although the characteristic of M-sequence is used to solve the problem of overlapping SCRs, it seems that it is necessary to use much more powerful methods for decomposing overlapping responses, e.g., such as the method proposed by Lim et al. [37].

Two types of classification methods were employed. Furthermore, all possible combinations of features were classified in order to examine their interactions. In both methods of classification, the best accuracy was achieved from the BCD and RLL. The incremental validity from the combination of measures was obtained using the LDA classification method, while no combination of the brain and autonomic measures yielded the incremental validity in logistic regression. The best results of different classification methods in different measures are summarized in Table 2. The absence of incremental validity in the

logistic regression method might be due to the type of this classification technique. Different methods of classification according to the assumptions and the rules used in their design might show different performances. This issue was observed in a review of classifiers in this study. Thus, the hypothesis of accuracy enhancement of the CIT combined with the ERP and autonomic measures can be validated.

Matsuda et al. [15] designed a CIT with simultaneous measurements of autonomic and brain signals. The cited authors used auditory stimulation with long ISIs (22 sec). The use of long ISIs causes no overlap between sequential autonomic responses, and these measures showed a significant difference between critical and noncritical items. The P300 amplitude did not show significant differences, which is probably due to the use of long ISIs and a low number of stimulus presentations. Gamer and Berti [16] performed a similar study sometime later. They designed a CIT with the combined measurement and ISIs of 7 to 9 sec and tried to improve the P300 amplitude discrimination using an increase in the frequency (and number) of stimulus presentations, but their results were similar to those reported by Matsuda et al. [15]. As was already mentioned, using short ISIs and considering some arrangements to solve the problems of autonomic measures, both the ERP and autonomic measures showed a significant difference in our study.

In a CIT study with multimodal measurements, Ambach et al. [11] used short ISIs (3.0-3.5 sec). As compared with similar studies, the cited authors reported that there is a significant difference between the probe and irrelevant items in the ERP and autonomic measures but with a smaller effect size. These researchers discussed the absence of a criminal context in the mock crime instruction and complete omission of answers, which led to the diminished involvement and attention of subjects as a possible reason for rather small effect sizes. They reported a 0.738 correct-classification rate by only P300 and a 0.829 correct-classification rate with combining P300 and SCR based on the logistic regression model between the guilty and hypothetical innocents. In our study, a higher effect size and a higher correct-classification rate were obtained. Most likely, the reason for this might be the subject's greater attention and stronger involvement in the experiment. The use

of target stimuli and the answering of all presented items by the subject (via mouse click) prevented reducing the subject's attention. Furthermore, the use of a criminal context in the mock crime instruction and the reward-punishment system (winning or losing the gold coin) have made the subjects more actively pay attention and get involved in the experiment.

Thus, our study showed that, both in the ERP and autonomic measures, a significant difference between the standardized difference scores of the guilty and innocent groups can be achieved using short ISIs and large numbers of stimulus presentations with the consideration of some arrangements. These arrangements include several blocks and variable ISIs to solve the problem of habituation, and also stimulus presentation in the balanced order (using M-sequence series) to solve the overlapping problem of the SCR. Furthermore, the criminal context in the mock crime instruction and reward-punishment system were used to make subjects to increase attention and get more involved in the experiment.

Finally, the hypothesis of accuracy enhancement of the CIT combined with the ERP and autonomic measures was confirmed by a review of two classification methods. To reach such a result, we tried to make experimental conditions closer to real conditions as much as possible using the criminal context in the mock crime instruction and real innocent subjects instead of hypothetical innocent ones.

Our own data show that the expediency of further studies is obvious. First, slightly longer ISIs should be used in order to elicit greater autonomic responses and to allow experimenters to use longer scoring intervals. Second, the method of decomposing the SCR in paradigms with short ISIs should be used [37, 38]. As a concluding suggestion, the use of more powerful classification methods, such as a support vector machine, might be useful for further studies.

All tested subjects were volunteers; they were informed in detail on the pattern of the experiment and gave informed consent.

The authors, E. D. Farahani and M. H. Moradi, declare that they have no conflict of interests.

Acknowledgment. This research was supported by the Iran National Science Foundation (INSF). The authors would like to thank them and also all those who took part in this study.

Е. Д. Фарахані¹, М. Мораді¹

ТЕСТ ІЗ ПРИХОВАНОЮ ІНФОРМАЦІЄЮ „ВИНУВАТИЙ/НЕВИНУВАТИЙ” У ПОЄДНАННІ З РЕЄСТРАЦІЄЮ ПОВ’ЯЗАНИХ З ПОДІЄЮ ПОТЕНЦІАЛІВ І ВИМІРЮВАННЯМ ВЕГЕТАТИВНИХ ПОКАЗНИКІВ

¹Технологічний університет Аміркабір, Тегеран (Іран).

Резюме

Тест із прихованою інформацією (ТПІ) базується на порівнянні фізіологічних реакцій суб’єкта в зондуючих і нейтральних ситуаціях, спрямованому на виявлення такої інформації. У нашій роботі ми визначали точність результатів ТПІ в умовах, коли цей тест поєднували з відведенням пов’язаних з подією потенціалів (ППП) і вимірюванням вегетативних показників. Особливістю нашого дослідження було максимальне наближення експериментальних умов до реальних за допомогою використання кримінального контексту в інструкції „макетування злочину” та „реально невинуватих” суб’єктів замість „гіпотетично невинуватих”. 52 волонтери виконували один із сценаріїв „винуватий/невинуватий”. ТПІ вміщував п’ять блоків з короткими міжстимульними інтервалами. У кожному з блоків стимули пред’являлись у балансованій послідовності сьомого порядку. Крім ЕЕГ-активності, реєстрували частоту пульсу, зміни шкірної провідності (ШП), дихальну активність і плетизмограму пальців. Статистичний аналіз показав, що між стандартизованими оцінками відмінностей характеристик як ППП, так і вегетативних показників у „винуватій” та „невинуватій” групах виявлялися вірогідні відмінності. При виявленні змін ШП очікувані результати, описані для стандартних результатів ТПІ, що базуються на вимірювання вегетативних показників, не досягалися. Порівняння двох класифікаційних методик показало, що поєднання відведення ППП і вегетативних вимірювань підвищує точність результатів ТПІ. Найбільша точність класифікації, отриманої із застосуванням лінійного дискримінантного аналізу, складала 90.9 %. Скоріш за все, використання кримінального контексту в інструкції „макетування злочину” і системи преміювання/покарання забезпечувало більший рівень уваги тестованих та їх більше залучення в експеримент, що й підвищувало точність тестування порівняно з такою в аналогічних дослідженнях.

REFERENCES

1. D. T. Lykken, *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, McGraw-Hill Book Comp., New York (1981).
2. J. A. Matte, *Forensic Psychophysiology Using the Polygraph: Scientific Truth Verification – Lie Detection*, J. A. M. Publ., New York (1996).
3. E. B. Ford, “Lie detection: Historical, neuropsychiatric and legal dimensions,” *Int. J. Law. Psychiat.*, **29**, 159-177 (2006).
4. National Research Council, *The Polygraph and Lie Detection*, Committee to Review the Scientific Evidence on the Polygraph, Division of Behavioral and Social Sciences and Education, The Nat. Acad. Press, Washington (2003).
5. V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, “A comparison of methods for ERP assessment in a P300-based GKT,” *Int. J. Psychophysiol.*, **62**, 309-320 (2006).
6. J. P. Rosenfeld, “Event-related potentials in the detection of deception, malingering, and false memories,” in: *Handbook of Polygraph Testing*, M. Kleiner (ed.), Acad. Press, New York (2002), pp. 265-286.
7. J. P. Rosenfeld, B. Cantwell, V. T. Nasman, et al., “A modified, event-related potential-based guilty knowledge test,” *Int. J. Neurosci.*, **42**, 157-161 (1988).
8. L. A. Farwell and E. Donchin, “The truth will out: interrogative polygraphy (lie detection) with event-related potentials,” *Psychophysiology*, **28**, 531-547 (1991).
9. D. T. Lykken, “The GSR in the detection of guilt,” *J. Appl. Psychol.*, **43**, 385-388 (1959).
10. G. Ben-Shakhar and E. Elaad, “The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review,” *J. Appl. Psychol.*, **88**, 131-151 (2003).
11. W. Ambach, R. Stark, M. Peper, and D. Vaitl, “A concealed information test with multimodal measurement,” *Int. J. Psychophysiol.*, **75**, 258-267 (2010).
12. J. Polich, “P300, probability, and interstimulus interval,” *Psychophysiology*, **27**, No. 4, 396-403 (1990).
13. R. J. Barry, S. Feldmann, E. Gordon, and K. I. Cocker, “Elicitation and habituation of the electrodermal orienting response in a short interstimulus interval paradigm,” *Int. J. Psychophysiol.*, **15**, 247-253 (1993).
14. E. H. Meijer, *Psychophysiology and the Detection of Deception: Promises and Perils*, Unpublished Doctoral Dissertation, Maastricht Univ., the Netherlands (2008).
15. I. Matsuda, H. Nittono, A. Hirota, et al., “Event-related brain potentials during the standard autonomic-based concealed information test,” *Int. J. Psychophysiol.*, **74**, 58-68 (2009).
16. M. Gamer and S. Berti, “Task relevance and recognition of concealed information have different influence on electrodermal activity and event-related brain potentials,” *Psychophysiology*, **47**, 355-364 (2010).
17. I. Matsuda, H. Nittono, and T. Ogawa, “Event-related potentials increase the discrimination performance of the autonomic-based concealed information test,” *Psychophysiology*, **48**, 1701-1710 (2011).
18. G. T. Buracas and G. M. Boynton, “Efficient design of event-related fMRI experiments using M-sequences,” *Neuroimage*, **16**, 801-813 (2002).
19. M. Soskins, J. P. Rosenfeld, and T. Niendam, “Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms,” *Int. J. Psychophysiol.*, **40**, 173-180 (2001).
20. M. E. Dawson, A. M. Schell, and D. L. Filion, “The electrodermal system,” in: *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (eds.), Cambridge Univ. Press (2000), pp. 200-223.
21. H. W. Timm, “Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique,” *J. Appl. Psychol.*, **67**, 391-400 (1982).
22. H. W. Timm, “Analyzing deception from respiration patterns,” *J. Police Sci. Adm.*, **10**, 47-51 (1982).

23. J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, **28**, R1-R39 (2007).
24. H. S. Shin, C. Lee, and M. Lee, "Adaptive threshold method for the peak detection of photoplethysmographic waveform," *Comput. Biol. Med.*, **39**, 1145-1152 (2009).
25. M. Velden and C. Wölk, "Depicting cardiac activity over real time: a proposal for standardization," *J. Psychophysiol.*, **1**, 173-175 (1987).
26. M. T. Bradley and M. P. Janisse, "Accuracy demonstrations, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures," *Psychophysiology*, **18**, 307-315 (1981).
27. G. Ben-Shakhar, "Standardization within individuals: a simple method to neutralize individual differences in skin conductance," *Psychophysiology*, **22**, 292-299 (1985).
28. M. Gamer, H. G. Rill, G. Vossel, and H. W. Gödert, "Psychophysiological and vocal measures in the detection of guilty knowledge," *Int. J. Psychophysiol.*, **60**, 76-87 (2006).
29. S. Wasserman and U. Bockenholt, "Bootstrapping: applications to psychophysiology," *Psychophysiology*, **26**, 208-221 (1989).
30. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Inc., Hillsdale (1988).
31. R. L. Rosnow and R. Rosenthal, "Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers," *Psychol. Methods*, **1**, 331-340 (1996).
32. D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, Hoboken (2000).
33. K. Fukunaga, *Statistical Pattern Recognition*, Acad. Press, New York (1990).
34. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Recognition*, Wiley-Interscience, New York (2001).
35. V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Comput. Methods Programs Biomed.*, **94**, No. 1, 48-57 (2009).
36. J. Allen, W. G. Iacono, and K. D. Danielson, "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 (1992).
37. C. L. Lim, C. Rennie, R. J. Barry, et al., "Decomposing skin conductance into tonic and phasic components," *Int. J. Psychophysiol.*, **25**, 97-109 (1997).
38. W. Ambach, R. Stark, M. Peper, and D. Vaitl, "Separating deceptive and orienting components in a Concealed Information Test," *Int. J. Psychophysiol.*, **70**, 95-104 (2008).