Biosignal-based Recognition Tests to Mitigate Insider Threat in Nuclear Facilities

Chul Min Kim [a], Hyeon-Jeong Suk[b], Man-Sung Yim [a, *]
[a] *Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology*
[b] *Department of Industrial Design, Korea Advanced Institute of Science and Technology*
[*] *Corresponding author: msyim@kaist.ac.kr*

1. Introduction

As safety/security-sensitive facility, nuclear power plants have paid a great attention to prevent human risk. However, the current fitness-for-duty evaluation and monitoring system relies on subjective and ex-post evaluations, such as pre-employment background check, mental health evaluation, and peer evaluation. As a result, the cost for safety/security has increased, at the expense of work efficiency. Therefore, the operators and regulatory agencies are demanding the technologies that can objectively evaluate the fitness-for-duty and reduce the burden of peer evaluation.

The biosignal-based approaches have shown the potential to solve this problem. Typical applications were fitness-for-duty management [1], human error reduction [2,3], and insider threat mitigation [4]. These technologies should contribute to reduce human risk by providing 1) real-time feedback on mental/physical conditions and mistakes of workers, 2) the degree of concentration of workers, and 3) communication between workers, thus reducing the burden of peer monitoring. Consequently, it is economical or safe from the human damage caused by human error or intentional error in facilities.

A previous study [5] attempted to classify the psychophysiological detection of deception. To mitigate insider threats it predicted the implicit intentions of participants on a given insider threat and daily stress scenarios. The scenario covers deception tests only, including comparison and non-comparison questions.

In this study, acquaintance test and concealed information test (CIT) were performed to test the technical feasibility of recognition tests to detect the deception. The results are expected to be supplementary indicators of the worker's trustworthiness. Furthermore, a biosignal-based insider threat mitigation policy can be established, in that both scenario-based malicious intention test and periodic recognition inspection will be properly planned.

2. Methods

We carried out acquaintance test and concealed information test using electroencephalography (EEG). Eight healthy young adults voluntarily participated in the experiment. Their EEG signals were recorded with 500 Hz, 21-channel international 10-20 system Ag/AgCl electrode cap and Neuron-spectrum 4/P (Neurosoft Ltd., Russia) software.

2.1 Acquaintance test

Acquaintance test is a task that aims to distinguish the biological reactions of seeing the photo of 20 unknown people, 4 acquaintances, and themselves (total 25 photos) using EEG. In single trial, a picture of a person appears for a period of 300ms and a black screen appears for a random period of time between 1500-1700ms. Each session consists of ten times of each stimulus presented, 250 trials in total, and the experiment was repeated over three sessions. Photos appeared in random order in each session. Participants were exposed to each stimulus on the monitor and instructed not to make any movement.

2.2 Concealed Information Test

In the CIT, the participants first committed a mock crime – on the desk in the experiment room, they picked up one out of five items (credit card, USB drive, wallet, key, SD card) presented on the table and put the selected item in his/her pocket. Then, when the participant's looking at the pictures of five items, biosignals were distinguished between the stolen item ("probe") and not stolen items ("irrelevant").

As shown in Fig. 1, the Complex Trial Protocol [6] proposed by Rosenfeld was implemented in this study. In this protocol, two sets of stimuli were presented for each trial. In the first trial, either probe or irrelevant was displayed for 300ms. Then, for a random time range between 1500-1700ms, participants pressed "a" button on the keyboard ("I saw it" response). In the second stimuli, target ("11111") and non-target ("22222", "33333", "44444", "55555") are displayed for 300ms. Additionally, for a random time range between 1500-1700ms, participants pressed the right arrow button if they saw the target, and pressed the left arrow button when they saw the non-target. The experiment was repeated over three sessions; each consisted of ten trials.

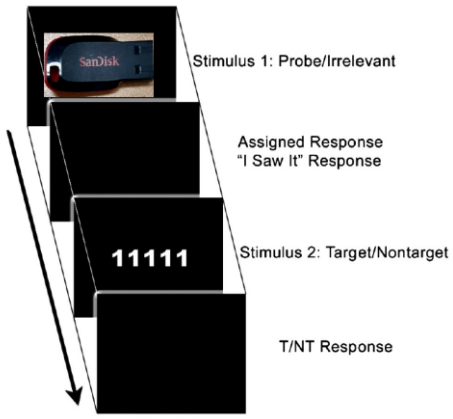

Fig. 1. Complex Trial Protocol for CIT [7].

2.3 Analysis of EEG Data

Preprocessing process was proceeded for the EEG data collected from 21-channels to remove noises and artifacts. The data were down-sampled to 250 Hz, and the frequency below 1 Hz was filtered using a high-pass filter. Besides, the external artifacts sich as body movements, eyeballs' rolling and eye blinking were discarded from the independent component analysis (ICA). The power at the electrode site Fz (frontal midline) were used for event-related potential (ERP) analysis.

Then stimulus-locked data were extracted for each trial, from -100 to 1500 ms ([-100, 1500] hereafter) relative to presentation of the stimulus. For artifact rejection, each trial was corrected to the baseline, a period of [-100,0]. Moreover, the trials with a range of greater than 100 uV between the highest and the lowest amplitude were removed from further analyses. Finally, the trials were averaged for each stimulus type to observe the difference of ERP shape among the stimulus types.

3. Results

3.1 Acquaintance Test

In acquaintance test, there was significant difference in the shape of ERP among the stimulus types (unknown, acquaintance and themselves). As shown in Fig. 2, the P300 ERP signal, defined by the difference between the highest and lowest power around 300ms, were the highest when the participants watched their own photos, followed by the acquaintances. The P300 ERP were the lowest in the photos of unknown people.

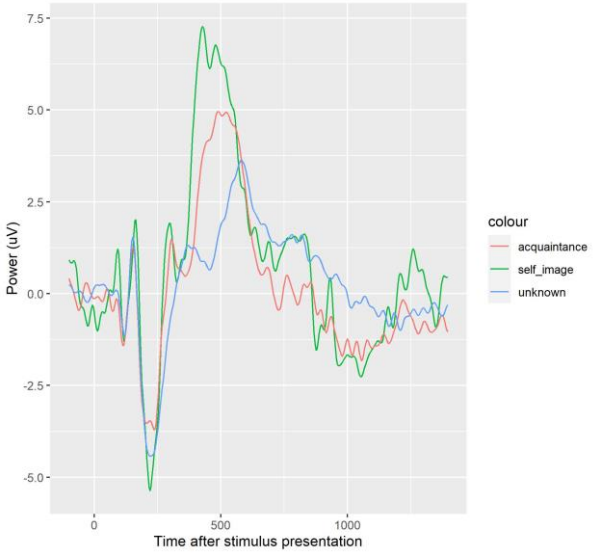

Fig. 2. Stimulus-locked ERP results in acquaintance test.

The results indicated the possibility of using P300 ERP for recognition tests for access control to vital areas in nuclear power plants, through the process of checking whether the worker does not recognize the specific person, possibly the terrorists, for example.

3.2 Concealed Information Test

Contrary to the previous CIT experiments, the P300 power was not significantly different between probe and irrelevant stimulus. Compared to the acquaintance test, the CIT's ERP appears to arise from the distinction between the unknown and the known stimuli, not from the intention to conceal specific information. In order for the CIT to properly generate the P300 ERP, the participant must be made unaware of the existence of the remaining four items. For example, it would be better for the participant to find 1 out of 5 hidden objects instead of choosing 1 out of 5 objects displayed on the desk.

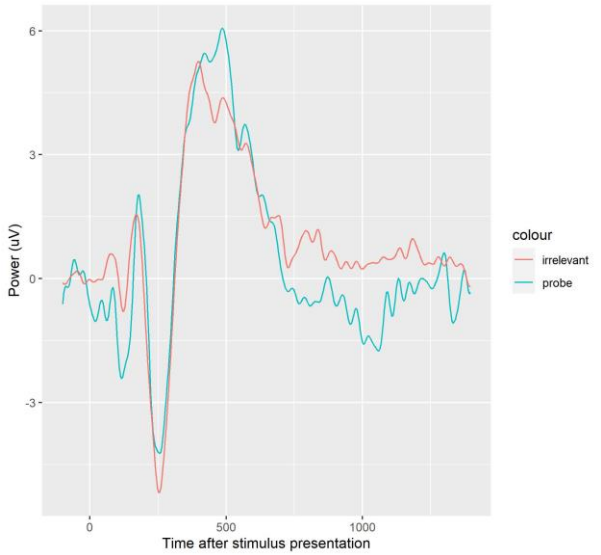

Fig. 2. Stimulus-locked ERP results in concealed information test.

4. Discussion and Conclusion

This study performed the biosignal-based recognition tests in two ways – acquaintance test and the CIT. The difference of P300 ERP was significant between the stimulus types in acquaintance test. On the contrary, the CIT failed to show the difference. It indicates that the use of CIT for normal workers should be limited to test the recognition and knowledge, rather than to test the intention of concealing any information.

Nonetheless, this study provides empirical evidence for the potential use of recognition tests for access control to vital area in nuclear power plants, under several critical considerations. First, although the P300 ERP distinguished the known people from the unknown, further investigation should be conducted to connect the recognition and the intention of deception to block the specific worker's access to vital areas. Second, other types of ERP need to be considered to identify the implicit intention in a comprehensive manner. For example, the real-time based error-related potential (ErrP) corrects the robot movements to detect malicious insiders.

Acknowledgments

This work was supported by the Nuclear Safety Research Program through the Korea Foundation of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea. (No. 2003023)

REFERENCES

[1] Y. A. Suh, and M.-S. Yim, A Worker's Fitness-for-Duty Status Identification Based on Biosignals to Reduce Human Error in Nuclear Power Plants, Nuclear Technology, 2020.

[2] J. H. Kim, C. M. Kim, Y. H. Lee, and M.–S. Yim, Electroencephalography-Based Intention Monitoring to Support Nuclear Operators' Communications for Safety-Relevant Tasks. Nuclear Technology, 2021.

[3] J. H. Kim, C. M. Kim, E. S. Jung, and M.–S. Yim, Biosignal-Based Attention Monitoring to Support Nuclear Operator Safety-Relevant Tasks, Frontiers in Computational Neuroscience, Vol. 14, No. 111, 2020.

[4] J. H. Kim, C. M. Kim, and M.–S. Yim, An Investigation of Insider Threat Mitigation based on EEG Signal Classification, Sensors, 20(21), 6365, 2020.

[5] D. J. Krapohl, J. B. McCloughan, and S. M. Senter, How to Use the Concealed Information Test, Polygraph, Vol. 35, No. 3, 2006.

[6] J. P. Rosenfeld, P300 in detecting concealed information and deception: A review, Psychophysiology, 2019.