Transactions of the Korean Nuclear Society Autumn Meeting
Goyang, Korea, October 24-25, 2019

Nuclear Power Plant Operator Performance Prediction using Facial Expression Analysis

Cho Woo Jo, Young Ho Chae, and Poong Hyun Seong*
Dept. Nuclear & Quantum Eng, Korea Advanced Institute of Science and Technology
291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
chowoo0701@kaist.ac.kr *Corresponding author: phseong1@kaist.ac.kr

1. Introduction

Until now, many nuclear power plant(NPP) accidents are caused by human error, and human error is directly related to safety of the NPP. There have been various studies [1-2] related to human error and accident diagnosis.

To reduce human error and predict performance, Radlo, S. J., et. al. [3] and Sun, J. C. Y., et. al. [4] suggested brain wave based human error prediction system. And also, Shagass, C., et. al. [5] proposed eye-tracking based performance system. Those gear wearing devices had strengths in providing real time data, however, there were worries if they might impair performance due to discomfort of wearing. In this regard, a wide range of applications of gear wearing devices in NPP was difficult. To overcome these limitations, we propose a practical method for predicting human performance in NPP.

Out of many bio signals, we found facial expression is the most prominent feature conveying real-time status information in an intrusive and invasive way. Also, unlike other bio signals, as facial expressions are associated with emotions, which influence information processing and perception [6], it was expected that facial expressions would imply human performance.

For these reasons, we at Nuclear Instrumentation & Control and Information Engineering Lab in KAIST chose facial expression analysis for nuclear operator performance prediction. There were several prior studies finding correlation between facial expressions and performance [7-8], but they were left with the possibility of further development as they focused only on general propensity. Facial expression analysis would provide real-time performance and feedback without affecting operator performance. From this research, we would find meaningful facial expression which might be connoted with important meanings.

To have a better knowledge of facial expression changes related to performance, we conducted human subject experiment and analyzed facial muscle movements. From the experiment, it was proved that certain facial muscle movements are related to performance. Further, it showed that facial expressions provide more precise performance prediction than conventional survey.

2. Methods

In this section, experiment sequences and some of the techniques used in experiment are described. Through human subject experiment, real-time facial expression data was analyzed, and the analyzed data was used to find correlation with operator performance. Then experiment results were compared with situation awareness and workload.

Experiment was done by 25 KAIST undergraduate and graduate students, who study science and engineering, as subjects.

First, subjects were trained nuclear safety system and characteristics of five major nuclear power plant accidents: Loss of coolant accident, Steam generator tube rupture, Loss of feed water accident, and Main steam line break (containment in and out). All experiment subjects were trained by the same recorded learning materials.

All the five accident situations were simulated in Compact Nuclear Simulator (CNS), made by Korea Atomic Energy Research Institute (KAERI) imitating Westinghouse PWR 930 MWe 3loop type.

The way to read simulator instrumentation values were taught in training session. After training, 30 seconds lasting CNS nuclear accident videos were provided to practice accident diagnosis.

When subjects were ready to start accident diagnosis, accident simulation videos were screened. Each five different videos were screened only one time. Then subjects were asked to diagnose accident based on their observation.

After experiment, SART (Situation Awareness rating technique) [9] and NASA-TLX (NASA-Task Load Index) questionnaire [10] were tested to measure subjects' situation awareness and workload.

Subjects were asked to act as if they are nuclear operators. For the experiment, cameras and commercialized facial expression analysis software, which were verified and validated by facial expression analysis experts, were used. [11-13]

Subjects were previously informed that videos would be taken. However, none of subjects have known their facial expression is analyzed until the experiment finished. Facial expressions were collected from the taken videos.

The 25 experiment subjects solved five questions in an accident diagnosis situation. Participants' performance was divided into two groups: good and poor performance.

3. Results

3.1 Facial Expressions and Human Performance

Independent sample t test was used to compare the result. The results were calculated by comparing average muscle movements per time.

From the experiment, we obtained interesting results that certain facial muscle moves often when human performance is bad. The result showed statistical difference in accordance with performance.

Among seven emotions (happiness, sadness, fear, surprise, anger, contempt, and disgust) and 20 facial muscle movements, we found one facial expression is meaningful to distinguish participants' performance: smirk. Smirk is one of facial muscles which represents a mug, conceited, or silly smile. Table I shows smirk facial expression difference with statistical importance regarding performance.

There were 125 performance results from 25 participants and their 5 answers. Among 125 answers, 98 of them were correct while the rest 27 were wrong. For facial expression results, however, some of data out of video screens were omitted as participants were not caught on the screen.

Table I: Comparison of smirk facial expressions average depending on performance.

	Average (μ)		Standard Deviation (σ)		t	P
	Low error	High Error	Low Error	High Error		
Low error = 92 High Error = 26	Smirk (0.757)	Smirk (5.254)	Smirk (3.752)	Smirk (12.66)	1.789	0.085

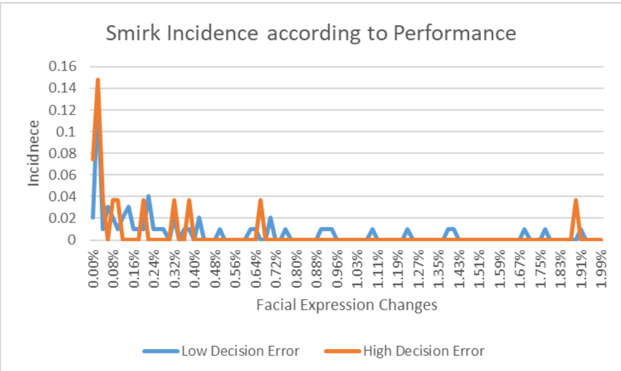

Fig. 1. Comparison of smirk facial expressions average depending on performance.

Table I represents comparison of smirk incidence in different performance situations. Average and standard deviation data represents feature of smirk depending on performance. According to Table I, average of smirk incidence is increasing with poor performance. It implies that people showing less smirk in a given task would record better performance.

In Fig. 1, Table I results were expressed to understand at a glance. To reduce the variance between different numbers of each group, we divided incidence by the number of good and poor performance group. Thus, Fig. 1 represents facial expression distribution for one people depending on performance. Blue line shows incidence distribution of smirk from participants with good performance, while orange line represents the other group's smirk incidence distribution. From the result, we found a group with poor performance have more distributed smirk movement. This is because bad performance group may or may not have anticipated their bad performance. In contrast, good performance group have more concentrated facial muscle movement distribution and also have smaller average. This results imply that good performance group were calm to diagnose accident as they had confidence in their decision.

This research showed significant difference in average smirk facial muscle movement according to performance. Based on the result, this research would proceed further by predicting operator real time performance with machine learning technique and also would specify certain error type to figure out which facial muscle movements are related.

3.2 Situation Awareness and Human Performance

The SART (Situation Awareness Rating Technique) questionnaire was conducted after experiment for assessing subjects' situation awareness (SA) in current experiment. Although situation awareness and human performance are normally assumed to be directly proportional, there have been research in 1988 done by Logan, saying that they are not [14]. This research focused on whether situation awareness can differentiate performance.

For experiment analysis, SART questionnaire answers were used to find correlation with performance. Crosstab analysis method was used. Table II represents situation awareness comparison depending on performance.

Table II: Comparison of situation awareness depending on performance

Unit: incidence (percent)

SA	Answer		Total	x^2	P
	Low error	High error			
1	0(0)	0(0)	0(0)	12.725	.003
2	3(3.1)	2(7.4)	5(4.0)		
3	0(0)	0(0)	0(0)		
4	34(34.7)	1(3.7)	35(28.0)		
5	47(48.0)	18(66.7)	65(52.0)		
6	14(14.3)	6(22.2)	20(16.0)		
7	0(0)	0(0)	0(0)		
Total	98(100)	27(100)	125(100)		

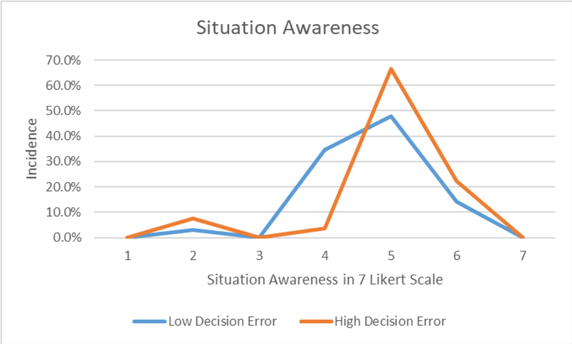

Fig. 2. Distribution: Comparison of situation awareness depending on performance.

From Fig. 2, situation awareness in 7-Likert scale did not show significant difference but showed similar trends according to performance. This results implies that situation awareness itself does not directly proportional to performance.

Interestingly, situation awareness with good performance was even inclined to lower level than situation awareness with bad performance. This tendency might have come from participants' self-evaluation, which means that people with high performance are tend to evaluate themselves strictly.

3.3 Workload and Human Performance

NASA-TLX survey was used to assess physical and mental perceived workload (WL) from a given task. In general, like situation awareness, people assume that high workload is highly related to low performance with more human errors.

Table Ⅲ shows perceived workload difference with performance. Although there is only a small amount of difference, it was found that workload with good performance is lower than with poor performance.

Table III. Comparison of workload depending on performance

Unit: incidence (percent)

WL	Answer		Total	x^2	p
	Low error	High error			
1	0(0.0)	0(0.0)	0(0.0)	8.646	.054
2	10(10.2)	0(0.0)	10(8.0)		
3	13(13.3)	2(7.4)	15(12.0)		
4	60(61.2)	15(55.6)	75(60.0)		
5	8(8.2)	7(25.9)	15(12.0)		
6	7(7.1)	3(11.1)	10(8.0)		
7	0(0.0)	0(0.0)	0(0.0)		
Total	98(100.0)	27(100.0)	125(100.0)		

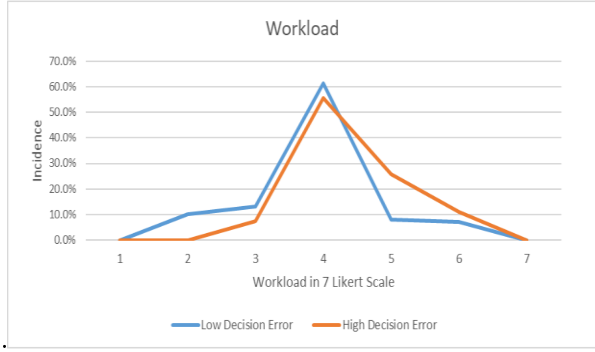

Fig. 3. Distribution: Comparison of perceived workload depending on performance.

From Fig. 3, it was found that workload was independent of performance but showed almost exact tendency. Also, most of participants assessed their workload at medium level.

4. Discussion

Considering facial expression, situation awareness, workload, we found there are statistical difference in smirk facial expressions according to performance. On the other hand, performance was hard to differentiate from conventional survey results.

From experimental results, situation awareness with good performance was differentiated from poor performance especially on fourth level of situation awareness, which means more people from good performance group would rate themselves to be less aware of situation. It was completely different from general thoughts that high situation awareness would lead high performance.

Meanwhile, perceived workload with good performance was also distinguished from poor performance on fifth level of perceived workload. This results represent that more people with poor performance felt more burden from the given task. This is consistent with the usual expectations.

Those two conventional survey results were available to differentiate performance, however, these are not enough to be used for performance prediction as the difference was relatively small and their overall changes according to situation awareness and workload scale were similar. Moreover, there were some limitations for conventional surveys that they do not provide a real time status changes, therefore we had to infer their real time status from survey results.

On the other hand, facial expressions analysis was more suitable to classify performance with more clear data separating boundaries. In the range of less than 0.08% of facial expression changes, more than three times of incidence was found in good performance group than poor performance group.

These are quite encouraging results as they can also say that human performance can be predicted real-time by facial expressions. As found from the results, it seems like by calculating average movements of certain facial muscle, we might predict their performance in the future.

5. Conclusions

As part of a plan to enhance nuclear safety, human error was the main consideration in nuclear power plant. In prior research, human error and performance prediction from bio signals has been tried, but it was hard to be applied in nuclear power plant because of practical issues. Thus, this research has focused on facial expression which is comparatively simple to observe status in real time.

It was expected that operator performance prediction using facial expression analysis can be a useful tool for predicting performance and reducing possible human error. Facial expression analysis has strengths in performance prediction as it provides real time data in an intrusive way.

From the experiment, we found facial expressions, especially smirk, shows different pattern according to performance. Thus, we concluded that facial expression would be useful to predict human performance. On the other hand, from conventional survey, situation awareness and workload do not show prominent difference according to performance.

In our future studies, this research would proceed the experiment with more people and further apply the results with machine learning technique to train model and identify possible human error. Besides, situation awareness, workload survey would be used supplementary indicators.

The ultimate goal of this research is to provide more reliable performance analysis software which would enhance nuclear safety. Therefore, we expect more meaningful facial expression changes from performance can be found from larger number of subjects. Moreover, we would like to subdivide human performance more precisely to identify certain error type.

Acknowledgments

This research was supported by the KUSTAR-KAIST corporation project through the Institute, KAIST, Korea funded by the KUSTAR university.
(N11190097)

REFERENCES

[1] Sheridan, T. B. (1981). Understanding human error and aiding human diagnostic behaviour in nuclear power plants. In Human detection and diagnosis of system failures (pp. 19-35). Springer, Boston, MA.
[2] Beare, A. N., & Dorris, R. E. (1983, October). A simulator-based study of human errors in nuclear power plant control room tasks. In Proceedings of the Human Factors Society Annual Meeting (Vol. 27, No. 2, pp. 170-174). Sage CA: Los Angeles, CA: SAGE Publications.
[3] Radlo, S. J., Steinberg, G. M., Singer, R. N., Barba, D. A., & Melnikov, A. (2002). The influence of an attentional focus strategy on alpha brain wave activity, heart rate and dart-throwing performance. International Journal of Sport Psychology, 33(2), 205-217.
[4] Sun, J. C. Y. (2014). Influence of polling technologies on student engagement: An analysis of student motivation, academic performance, and brainwave data. Computers & Education, 72, 80-89.
[5] Shagass, C., Roemer, R. A., & Amadeo, M. (1976). Eye-tracking performance and engagement of attention. Archives of General Psychiatry, 33(1), 121-125.
[6] Forgas, J. P., & Bower, G. H. (1987). Mood effects on person-perception judgments. Journal of personality and social psychology, 53(1), 53.
[7] Stone, R. T., & Wei, C. S. (2011, September). Exploring the linkage between facial expression and mental workload for arithmetic tasks. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 55, No. 1, pp. 616-619). Sage CA: Los Angeles, CA: SAGE Publications.
[8] Whitehill, J., Bartlett, M., & Movellan, J. (2008, June). Automatic facial expression recognition for intelligent tutoring systems. In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (pp. 1-6). IEEE.
[9] Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness (pp. 111-128). Routledge.
[10] Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.
[11] Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. Environmental psychology and nonverbal behavior, 1(1), 56-75.
[12] Senechal, T., McDuff, D., & Kaliouby, R. (2015). Facial action unit detection using active learning and an efficient non-linear kernel approximation. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 10-18).
[13] Ekman, R. (1997). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA.
[14] Logan, G. D. (1988). Toward an instance theory of automatization. Psychological review, 95(4), 492.