Toward a novel situation assessment (SA) measure*

Jinkyun Park[a], Yochan Kim[a], and Wondea Jung[a]
[a] Korea Atomic Energy Research Institute, Daejeon, Republic of Korea

Abstract: It is well recognized that one of the significant factors affecting the performance of human operators is the level of situation awareness (or situation assessment, SA). This means that the evaluation of SA is very important for clarifying the variation (either enhancement or deterioration) of human performance under diverse task environments, which is critical for the safe operation of socio-technical systems, such as nuclear power plants (NPPs). For this reason, many researchers have proposed various kinds of methods and/or measures for visualizing the level of SA. Unfortunately, existing methods and/or measures have common limitations. For this reason, a novel SA measure is proposed by the collaboration of the process mining technique and signal detection theory (SDT). In addition, SA scores estimated by the proposed measure are compared with those came from the application of an existing SA measure called the SART (Situation Awareness Rating Technique). As a result, it is observed that there is a significant correlation between the estimated SA scores and the associated SART scores.

Keywords: Situation assessment, Nuclear power plant, Signal detection theory, Process mining technique, Objective measurement.

1. INTRODUCTION

According to an existing study, it is revealed that 15 issues are important for understanding the performance of human operators who are working in either analog or digital environments. Table 1 summarizes 15 issues with respect to nine categories [1].

Table 1: 15 human performance issues belonging to nine categories (adopted from Ref.[1])

Category	No.	Detailed human performance issue
HMI complexity	1	Misplaced salience
Situation awareness (SA)	2	Keyhole effect
	3	Out-of-the-loop with the level of automation
	4	Lack of early detection
	5	Missing task critical information
	6	Requisite memory trap
Cognitive workload	7	Cognitive workload due to alarm overload
	8	Cognitive workload due to excessive nuisance alarms
	9	Cognitive workload due to data overload
Physical workload	10	Physical workload
Crew performance	11	Coping with complex disturbances
Opacity in a digital system	12	Complexity creep
Dealing with diverse information across different sources	13	Concurrent use of analog and digital systems
Fatigue due to digital environment	14	Anxiety, time pressure, work criticality, and other stressors
Confirmation/trust on a (digital) system	15	Low trust in sensor readings

* More detailed background and technical basis about this paper can be found from: Ronald L. Boring, Thomas Ulrich, and Bruce P. Hallbert, Jinkyun Park, Yochan Kim, and Wondea Jung (KAERI), INL/EXT-17-43719 (KAERI/TR-6968/2017), Evaluation of the sustainability and effectiveness of proposed methods and measures for operator performance in control rooms (2017).

As can be seen from Table 1, one of the critical issues is the situation awareness (or situation assessment, SA) of human operators. Although there are a couple of SA definitions, the most popular one would be proposed by Endsley [2]: "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near feature." Based on this model, Endsley suggested a conceptual model which consists of three SA levels (i.e., *Level 1*, *Level 2*, and *Level 3* in Fig 1). In short, SA *Level 1* denotes the ability of human operators to perceive and/or detect significant information, while SA *Level 2* implies their competence to capture the nature (or feature) of an on-going situation by integrating and/or translating remarkable information perceived from SA *Level 1*. Once the nature of an on-gong situation was successfully identified, it is possible for human operators to come up with proper responses (or countermeasures) through predicting the progression of a situation at hand (i.e., SA *Level 3*).

In this light, it is expected that the SA impairment of human operators could the main source of diverse human performance issues. For example, a poor SA *Level 1* results in the loss of critical information or a lack of early detection, which trigger the degradation of human performance. Similarly, it is reasonable to assume that a keyhole effect (i.e., focusing on several prominent symptoms instead of considering various kinds of available information) is apt to distort the whole picture of an on-going situation, which can be explained as the result of a poor SA *Level 2*.

For this reason, many researchers have spent their effort for many decades in order to systematically evaluate and/or measure the SA of human operators. In this regard, Uhlarik and Comerford distinguished existing SA measures and/or methods into three groups with six subcategories as summarized in Table 2 [3].

Table 2: Existing SA measures and/or methods (adopted from Ref. [1])

Group	Subcategory	Remark
Explicit measure	Retrospective measure	Based on recall of specific events or the description of decisions made during experiments
	Concurrent measure	On-line measurements by using verbal protocols
	Freeze technique	Questionnaires collected from several points that were randomly frozen during experiments
Implicit measure	Global measure	Predicting the level of SA by using overall task performance
	External task measure	Correlating the performance of human operators with the level of SA based on the observation from the situation in which information is removed or the alternation of information is displayed
	Embedded task measure	Correlating the performance of human operators who have to carry out several sub-tasks with the level of SA
Subjective measure	Direct self-rating	Using self-rated questionnaires for measuring the level of SA (e.g., Situation Awareness Rating Technique, SART)
	Comparative self-rating	Requiring participants to compare self-assessed SA from one trial to another
	Observer rating	Requiring unbiased and neutral experts who rate the participant's SA level

However, it seems that existing SA measures and/or methods include one of the following limitations, such as subjectivity, resource intensive work, and intervening human operators. In order to clarify this claim, let us consider SART (Situation Awareness Rating Technique), which is one of the representative methods for quantifying the SA of human operators. Basically, the SART measures the SA of human operators by analysing their self-rating scores (7-point Likert scale) with respect to ten questions summarized in Table 3.

Table 3: Ten questions for the SART (adopted from Ref. [4])

ID	Dimension	Description
1	Instability of	How changeable is the situation? Is the situation highly unstable and likely to

	situation	change suddenly (High) or is it very stable and straightforward (Low)?
2	Variability of situation	How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing (Low)?
3	Complexity of situation	How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?
4	Arousal	How around are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?
5	Spare mental capacity	How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?
6	Concentration	How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focused on only one (Low)?
7	Division of attention	How much is your attention divided in the situation? Are you concentrating many aspect of the situation (High) or focused on only one (Low)?
8	Information quantity	How much information have you gained about this situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?
9	Information quality	How much information have you gained about this situation? Have you received high degree of goodness of knowledge (High) or do you have a low degree of goodness (Low)?
10	Familiarity	How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?

Based on the self-rating scores of 10 questions, the SA of human operators can be quantified using the following formulas [5]:

SA = Understanding – (Demand – Supply), where

Understanding = sum of subjective ratings for the questions of 8, 9, and 10;

Demand = sum of subjective ratings for the questions of 1, 2, and 3;

Supply = sum of subjective ratings for the questions of 4, 5, 6, and 7

Unfortunately, in order to use the SART, it is indispensable to resolve several problems, such as the subjectivity of self-rating scores and frequently asking 10 questions to human operators (i.e., resource intensive works and intervening them). In addition, the translation of SART scores needs the comprehension of each question (i.e., high expertise). Accordingly, it is necessary to develop a novel SA measure which can soundly resolve the abovementioned limitations. To this end, in this study, a novel SA measure is proposed through analyzing detailed action logs with SDT (Signal Detection Theory). It should be noted that a process mining technique was used as a tool for extracting key features from action logs, which specifies the whole history of information navigation done by human operators. More detailed explanations are given in the following sections.

2. EXTRACTING KEY FEATURES FROM ACTION LOGS

In terms of training for human operators working in socio-technical systems such as NPPs, it is inevitable to use a full- or partial-scope simulator which allows them to experience the situation of a very rare event (e.g., off-normal events). For example, Figure 1 shows the layout of a full-scope simulator which is the replica of a main control room installed in APR1400 (Advanced Power Reactor 1400MWe). Actually, since the primary purpose of this full-scope simulator is to train professional operators working in the main control room of ARP1400, one of the basic functions being implemented is to store all kinds of actions in the form of a text log. Typical example is shown in Fig. 2 which contains a part of an action log. For example, the first line of Fig. 2 informed us that a reactor operator (RO) changed the third screen in his and/or her workstation in order to pop up an information display page which is related to a pressurizer (i.e., “PZR____9_431_j168_107_02”). This implies that we are able to identify the whole history of information navigation done by the RO through the analysis of the corresponding action log.

In this regard, Park et al. [7] claimed that a process mining technique can be regarded as an appropriate tool for extracting key features including the history of information navigation from action logs. According to van der Aalst, the process mining technique have three kinds of basic competences: “The first type of process mining is discovery. A discovery technique takes an event log and produces a model without using any a-priori information. Process discovery is the most prominent process mining technique. For many organizations it is surprising to see that existing techniques are indeed able to discover real processes merely based on example executions in event logs. The second type of process mining is conformance. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is enhancement. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model [8].”

Figure 1: Layout of a full-scope simulator for APR1400 (adopted from Ref. [6])

```
2015/11/07 15:40:29:0679 => (ro3) Screen Change
                                   (DRAWING_ID:PZR____9_431_j168_107_02,
                                   EXTENDED_INFO:L)
                                               ...
2015/11/07 15:43:40:0539 => (ro1) Screen Change
                                   (DRAWING_ID:DE_____9_481_j168_101_04,
                                   EXTENDED_INFO:L)
                                               ...
```

Figure 2: A part of an action log (adopted from Ref. [7])

Actually, if we focus on the first competence of the process mining technique (i.e., discovery), it is possible to soundly extract key features from action logs. For example, Fig. 3 depicts a part of a graph which denotes the interrelation of display pages visited by human operators. From Fig. 3, the catalog of information display pages which were actually visited by human operators (e.g., CNDSR, RCFC, and RCP2A).

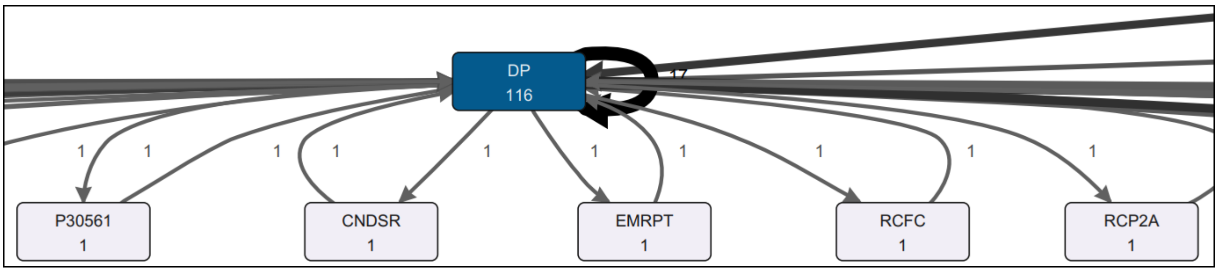
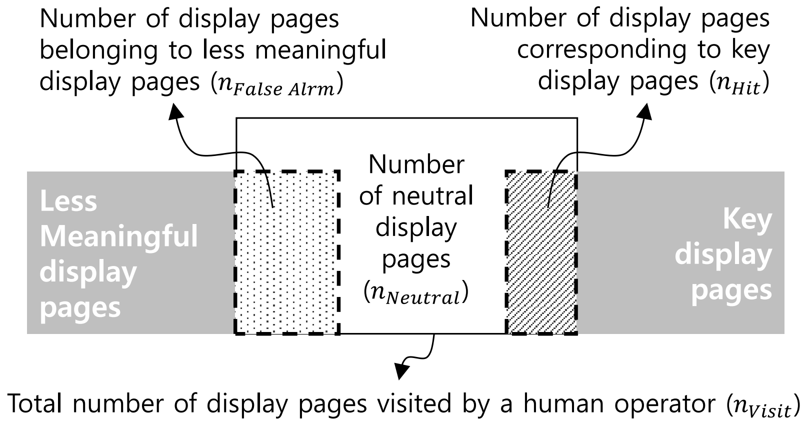


Figure 3: A part of a graph containing the interrelation of information display pages

3. APPLYING SDT

If we are able to identify the catalog of information sources observed by human operators (e.g., information display pages), it is likely that proper measures for dealing with the SA of human operators can be proposed. For example, let us assume that the information display pages of *RCFC* and *RCP2A* in Fig. 3 contain a couple of symptoms which play a critical role in detecting what has happened (i.e., *Level 1* SA). In this case, if there are two human operators who have different information navigation history, it is possible to distinguish their SA levels based on the catalog of key information display pages. Figure 4 shows this underlying idea.



$$n_{Visit} = n_{Hit} + n_{False\ Alarm} + n_{Neutral}$$

$$p(False\ Alarm) = \frac{n_{False\ Alarm}}{n_{Visit}} \qquad p(Hit) = \frac{n_{Hit}}{n_{Visit}}$$

Figure 4: A part of a graph containing the interrelation of information display pages (adopted from Ref. [9])

As can be seen from Fig. 4, the catalog of information display pages which human operators actually visited (or pop up) can be reclassified into three groups: (1) key display pages containing important information and/or symptoms for the performance of a required task, (2) neutral display pages that provide task neutral information (e.g., directory pages or common information display pages), and (3) less meaningful display pages that are not directly related to the performance of the required task. Once these three groups are properly distinguished, metrics from the signal detection theory (SDT) can be applied to the measurement of SA level.

Although the SDT has been used in the field of psycho-physics, it was proposed in order to analyse the performance of communication systems. Basically, the SDT deals with a problem, in which there are two kinds of discrete inputs (e.g., *Signal* and *Noise*) and the corresponding outputs (e.g., *Yes* and *No*). In this condition, the SDT suggests diverse measures based on four kinds of conditional probabilities (i.e., *Hit*, *Miss*, *False Alarm*, and *Correct Rejection*), which are also applicable for measuring the performance of human operators (Adbi, 2009). One of the representative measures is the sensitivity of the sensory process, which is computed by the following formula [10].

$$\text{Sensitivity} = \frac{p(Hit) - p(False\ Alarm)}{1 - p(Hit)}$$

Here, it is very interesting that this sensitivity measure can be used to quantify the SA of a human operator. For example, it is possible to assume that when the *Level 1* SA of a human operator is low, he or she will try to specify the most probable hypothesis by navigating diverse information display pages with low sensitivity. In contrast, if *Level 1* SA is high, he or she will focus on several information display pages which contain obvious symptoms for supporting the hypothesis in mind.

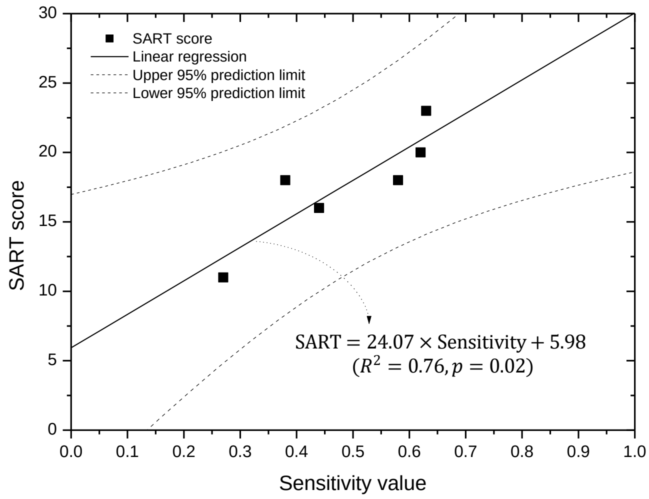
This means that, in general, the sensitivity of human operators who have high *Level 1* SA would be higher than those who have a low *Level 1* SA.

4. CASE STUDY

As explained at the end of the last section, it is expected that the sensitivity score of human operators would increase along with the increase of *Level 1* SA. In order to investigate this hypothesis, the sensitivity scores of six human operators were compared with their SART scores. All of the human operators are male and working in the main control room of APR1400. SART scores were collected from all human operators when they finished a simulated SGTR scenario. Table 4 summarizes sensitivity scores and the associated SART scores. In addition, Fig. 5 shows the result of comparisons between sensitivity scores and the associated SART scores.

Table 4: Comparing SART scores with the associated sensitivity scores (Revised from Ref. [9])

Operator ID	SART score	Sensitivity value
1	11	0.27
2	18	0.38
3	20	0.62
4	16	0.44
5	18	0.58
6	23	0.63



ANOVA (Analysis Of Variance) result

$$\text{SART} = 24.12 \times \text{Sensitivity} + 5.93\ (R^2 = 0.77, p = 0.02)$$

Item	Degree of freedom	Sum of squares	Mean square	F statistics
Model	1	62.54	62.54	13.31
Error	4	18.79	4.70	-
Total	5	81.33	-	-

Figure 5: Results of statistical analyses with respect to Table 4 (Adopted from Ref. [9])

From Fig. 5, it is evident that sensitivity scores are proportional to proportional to the increase of SART scores. This strongly implies that the sensitivity scores would be useful for representing the *Level 1* SA of human operators. Actually, this expectation can be supported by the results of statistical analyses, such as R^2 (i.e., 77% of variations related to the SART scores can be explained by those of sensitivity values) and F statistics (i.e., $p < 0.05$).

5. DISCUSSION AND CONCLUSION

As can be seen from Fig. 5, it seems that the sensitivity score is a good measure to quantify the SA of human operators (at least *Level 1* SA). If so, the use of the sensitivity score is beneficial because it allows us to resolve the common limitations of existing SA measures and/or methods, such as (1) a lack of objectivity (e.g., most of SA measurements depend on the subjective rating of human operators), (2) low usability (e.g., difficult to estimate the level of SA without intervening of human operators), (3) high effort (e.g., a series of questions should be asked to measure the SA of human operators), and (4) high expertise (e.g., highly experienced evaluators are needed to measure the SA of human operators).

It is still careful to confirm this expectation based on the results of this study because of a small number of human operators with a single off-normal scenario (i.e., an SGTR). Therefore, it is indispensable to conduct further analyses with more off-normal scenarios and human operators. However, the use of the sensitivity score proposed in this study is still beneficial because it provides a technical underpinning related to the quantification of diverse SA levels. That is, sensitivity sores given in Table 4 are representative for *Level 1* SA because the catalog of key information display pages considered in this study is preliminary limited to those related to information detection and/or gathering. In contrast, if we identify the catalog of key information display pages which are critical for understating the nature of a situation at hand, the sensitivity score of each operator (i.e., *Level 2* SA) can be soundly quantified by using the same concept depicted in Fig. 4. In this light, this study would be a good starting point to come up with further research directions for measuring the SA of human operators.

Acknowledgements

This work was supported by Nuclear Research & Development Program of the National Research Foundation of Korea (NRF) grant, funded by the Korean government, Ministry of Science and ICT (Grant Code: 2014M2A8A2074098).

References

[1] J. Park, W. Jung, R. L. Boring, T. Ulrich, C. Kovesdi, K. Savchenko, and B. P. Hallbert. Human performance issues in analog, hybrid, and digital control rooms, INL/EXT-15-37277, Idaho National Laboratory (2015).
[2] M. R. Endsley. Toward a theory of situation awareness in dynamic systems, Human Factors, 37, pp.32-64 (1995).
[3] J. Uhlarik, and D. A. Comerford. A review of situation awareness literature relevant to pilot surveillance functions, DOT/FAA/AM-02/3, Federal Aviation Administration (2002)
[4] Eurocontrol. https://ext.eurocontrol.int/ehp/?1=node/1608 (2017).
[5] S. J. Selcon and R. M. Taylor. Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In: Advisory Group for Aerospace Research and Development (AGARD), Conference Proceedings No. 478, p. 5-1 – 5-8 (1989).
[6] KNP news. http://www.knpnews.com/news/quickViewArticleView.html?idxno=1749 (2015).
[7] J. Park, J. Jung, G. Heo, Y. Kim, J. Kim, and J. Cho. Application of a process mining technique to identifying information navigation characteristics of human operators working in a digital main control room – feasibility study, Reliability Engineering and System Safety, vol. 175, p. 38-50 (2018).
[8] G. Vossen. The Process Mining Manifesto - An interview with van der Aalst, Information Systems, vol. 37, p. 288–290 (2012)
[9] R. L. Boring, T. Ulrich, B. P. Hallbert, J. Park, and W. Jung. Evaluation of the sustainability and effectiveness of proposed methods and measures for operator performance in control rooms, INL/EXT-17-43719 (KAERI/TR-6968/2017), Idaho National Laboratory (2017)
[10] H. Abdi. Signal detection theory. In: Encyclopedia of Education (McGaw, B., Perterson, P. L., and Baker, E., Eds.), 3rd Ed., Elsevier, New York (2009)