Nuclear Engineering and Technology 48 (2016) 129–143

Available online at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Method for Inference of Operators' Thoughts from Eye Movement Data in Nuclear Power Plants

Jun Su Ha [a,*], *Young-Ji Byon* [a], *Joonsang Baek* [a], *and Poong Hyun Seong* [b]

[a] *Khalifa University of Science, Technology & Research, P.O. Box 127788, Abu Dhabi, United Arab Emirates*
[b] *Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea*

A R T I C L E I N F O

Article history:
Received 2 July 2015
Received in revised form
7 September 2015
Accepted 9 September 2015
Available online 27 October 2015

Keywords:
Diagnosis
Eye tracking
Monitoring
Nuclear power plant
Selective attention

A B S T R A C T

Sometimes, we need or try to figure out somebody's thoughts from his or her behaviors such as eye movement, facial expression, gestures, and motions. In safety-critical and complex systems such as nuclear power plants, the inference of operators' thoughts (understanding or diagnosis of a current situation) might provide a lot of opportunities for useful applications, such as development of an improved operator training program, a new type of operator support system, and human performance measures for human factor validation. In this experimental study, a novel method for inference of an operator's thoughts from his or her eye movement data is proposed and evaluated with a nuclear power plant simulator. In the experiments, about 80% of operators' thoughts can be inferred correctly using the proposed method.

1. Introduction

Sometimes, we need to figure out somebody's thoughts from his or her behaviors such as eye movement, facial expression, gestures, and motions. Usually, inference of somebody's thoughts from his or her behaviors is associated with a lot of uncertainty, because the same behaviors might have different meanings depending on the context. However, if a person is carrying out his or her job in a very specific situation, the uncertainty coupled with the inference of the person's thoughts from his or her behaviors can be reduced and the inference of thoughts could be utilized for some useful applications. Operational tasks in control rooms of nuclear power plants (NPPs) are one of the representative examples that have very specific job characteristics. Generally, operators' tasks in NPPs constitute cognitive activities such as monitoring and detecting the environment; understanding, assessing, and diagnosing situations; decision making; planning responses; and implementing responses [1]. If operators' thoughts on a situation (or diagnosis result) can be inferred from the observation of their behaviors, this knowledge would have great potential for enhancing safety during NPP operation. As shown in the Tree Mile Island accident, a correct diagnosis has been considered as one of the most critical contributions to safe operation of NPPs

* *Corresponding author.*
E-mail address: junsu.ha@kustar.ac.ae (J.S. Ha).

http://dx.doi.org/10.1016/j.net.2015.09.004

[2]. As digitalized instrumentation and control systems have rapidly been applied to various plant systems including NPPs, operators' role in advanced control rooms has been changed from a controller to a supervisor [3–5]. In the majority of cases, the primary means of information input to operators in NPP control rooms is through the visual channel. Operators in NPPs are required to monitor several information sources such as indicators, alarms, controllers, and mimic displays provided in advanced control rooms, but they have limited capacity with respect to attention and memory. Hence, NPP operators pay selective attention to relevant and important information sources to effectively understand the current status [6].

Yarbus [7] conducted an important eye tracking research, showing that the task given to a person has a very large influence on that person's eye movement. The eye movement pattern during examination of pictures was dependent not only on what was shown in the picture, but also on the problem faced by the observer and the information that he or she hoped to gain from the picture. Eye movement data have also been studied in the field of intention inference. A fuzzy inference method was proposed to infer human intentions from eye movement data during monitoring of tasks on a computer screen [8]. The intention of the observer to move eye gaze from one point to another on the screen was inferred by applying a fuzzy logic based on a fuzzy set. A truck driver's intention to change lanes could be detected and inferred from his or her eye movement data [9]. The driver's eye movement pattern while changing the lane was analyzed, and a state transition model representing the likelihood of the driver changing the lane was developed based on the analyzed pattern. The authors have developed measures of fixation to importance ratio (FIR) and selective attention effectiveness (SAE) based on eye movement data, which represent how effectively an operator attends to important information sources as measures of attentional-resource effectiveness in monitoring tasks [10]. The FIR is the ratio of attentional resources (i.e., the number and duration of eye fixations) used on an information source to the importance of the information source. The SAE incorporates the FIRs for all information sources. In this experimental study, a novel method for inference of the operators' thoughts (understanding or diagnosis of a current situation) from their eye movement data is proposed and evaluated. The inference method is based on the FIR and SAE evaluation. In the second section, the cognitive processes of attention, understanding, and diagnosis are addressed to understand the principles underlying the method of approach, and the FIR and the SAE are introduced. The inference method is explained with examples of accidents occurring in NPPs. The experimental study is addressed in the third section followed by discussion in the fourth section, and the conclusion in the fifth section.

2. Inferring operators' thoughts from eye movement data

2.1. *Attention, understanding, and diagnosis in NPPs*

NPP operators generally keep monitoring the plant systems to detect any problems that may take place during normal operation. If an abnormal situation is detected, they search for relevant information sources to figure out what the problem is. Information processing is dependent on a pool of attention or mental effort, which is of limited availability and can be allocated to processes as required [11]. In terms of attentional resources, selection of information sources for further information processing should be addressed, as well as dividing attention between tasks. Selection of information sources is governed by salience, expectancy, value, and effort [12]. Salience refers to stimuli in the environment such as alarms, alerts, or some remarkable indication prompting attention. Expectancy makes attention shift to specific sources that are most likely to provide information. The value of the information source adjusts the frequency of looking at it. If too much effort is required in comparison with the value of the information source, attention might be restricted. The first studies on monitoring or information searching behavior were carried out for flight maneuver tasks in the late 1940s and early 1950s [13–15]. The relative importance (value) of information sources was reported as a governing factor in information searching behavior during flight maneuver tasks [15]. Senders showed that bandwidth (event rate) also plays a significant role in monitoring tasks [16], which have been subsequently elaborated to consider value with the bandwidth by a lot of researchers [17–40]. The bandwidth contributes not only to expectancy, but also to value. It provides operators with the expectancy of the location of information sources and valuable information for diagnosis in more detail. For example, if there is a rupture of a pipe through which water flows, the change in the flow rate of water due to the rupture provides information on the size of the rupture. Effort and salience may influence the selection of information sources to the extent that designers have adhered to good human factor practice in display layout [26]. Hence, effort and salience should be considered during the design phase by correlating effort and salience with expectancy and value.

When multiple tasks require to be performed at the same time, a strategy must be developed for dividing attention or allocating resources between tasks [41,42]. Perception or understanding consists of three simultaneous processes: bottom-up processing, top-down processing, and unitization (or matching). Stimuli or salient information sources derive the bottom-up processing through sensing mechanisms. After detecting a stimulus, the information is matched to a mental model that is established based on knowledge and experience. The effective selection of information sources is made by expectancy derived from the mental model, which is referred to as top-down processing. The chain of bottom-up processing, top-down processing, and unitization is the process of perception or understanding. A lot of information sources are provided in NPP control rooms, and operators have limited capacity of attention and memory. Operators have to selectively allocate their attentional resources. Selective attention is employed to overcome the limitations of human attention, making use of both top-down and bottom-up processes [12]. In abnormal situations in an NPP, operators collect a bunch of information from the human–machine interface (HMI) or other operators, and try to understand the abnormal situation, which is a process of establishing a situation model based on their mental model. The situation model is constantly updated as new information is received [43]. The mental model represents the general knowledge of a whole system. In NPPs, the mental model refers to the knowledge of plant system behaviors in various situations, which derive expectancies about information sources during abnormal situations. Knowledge of system behaviors in NPPs should be established in an operator's mental model through education, training, and experience. For example, if a loss of coolant accident (LOCA) occurs in an NPP, pressure, temperature, and level of the pressurizer (PRZ) would decrease, and the containment radiation would increase. These expectancies form some rules about plant system behaviors, and operators' mental models are established based on these rules. Usually, alarms or indicators that show deviation from normal conditions help operators detect a problem. The operators then develop their situation model with selective attention to important information sources. Moreover, an update of their situation model is achieved by iterating the selective attention. Hence, effective selective attention should correspond to correct understanding.

In highly complex technical installations involving high hazards such as NPPs, diagnosis is a crucial part of disturbance control [44]. A diagnostic task for response selection during abnormal situations in NPPs generally involves monitoring and detection of an abnormal situation, investigation of symptoms, reasoning for a possible cause, and a diagnostic judgment of the current situation. As a representative example of accidental situations, the steam generator tube rupture (SGTR) is investigated to understand the diagnosis process in NPPs. If the pressure, temperature, and level of the PRZ decrease, then an SGTR and an LOCA would be competing hypotheses. In order to diagnose the accident correctly, other parameters should be checked out. The LOCA is usually associated with an increase in the containment radiation. If there is no change in the containment radiation, the accident would be an SGTR. A set of symptoms is usually associated with an abnormal state of an NPP (i.e., situation–events relations) [37]. Considering that symptoms usually provide information for diagnosis of an accident, two kinds of symptoms should be addressed: symptoms representing a changed part (e.g., onset of alarm or deviation in a process variable) and symptoms that are not changed but provide diagnostic information. In the STGR example, no change in the containment radiation represents a kind of stationary symptom capable of differentiating the SGTR from the LOCA. Even when the symptoms are not changed, operators should pay selective attention to stationary symptoms so that they understand the situation correctly. The expectancy on these symptoms plays a significant role for correct diagnosis. After figuring out what the accident is, additional diagnoses might be required to find out details such as the location and size of a rupture, the amount of coolant leaked, and so on, which can be investigated by analyzing the bandwidths (event rates) of the process parameters (symptoms). If the process parameters of symptoms change very fast, operators are supposed to diagnose the situation as a large SGTR. On the contrary, if their values change slowly, operators would assess the situation as a small SGTR. Hence, the correct diagnosis during an abnormal situation should correspond to effective selective attention on important (or relevant) symptoms. Sets of behaviors (dynamics) of the simulator used in this experimental study are summarized in Table 1.

2.2. *Attentional-resource effectiveness*

Attentional resources of operators in NPPs should be allocated to valuable sources of information in order to effectively monitor, detect, correctly understand, and diagnose the state of a system, since operators receive too much information and they have limited attentional resources. Two measures of attentional-resource effectiveness have been developed by the authors for monitoring tasks in NPPs [10]. The measures are based on the underlying principle that attentional resources should selectively be allocated to information sources according to their informational importance. In the SGTR example, the pressure, temperature, and level of the PRZ are definitely important information sources, and selective attention should be paid to these sources in proportion to their importance. A measure of attentional-resource effectiveness for an individual information source is defined as the relative attentional resources consumed on the information source divided by the relative importance of the information source. The attentional resource to importance of information source ratio (AIR) is given as follows:

$$AIR(i) = \frac{\text{relative attentional} - \text{resources on information } - i}{\text{relative importance of information } - i} \tag{1}$$

Both the relative attentional resources and the relative importance of each information source should be normalized to range from zero to unity because they are relative measures. Information is provided to operators in NPP control rooms primarily through the visual channel. Hence, the AIR is converted into a measure that is expressed in terms of visual resources, FIR, as follows:

$$FIR^{N}(i) = \frac{\frac{N_i}{\sum_{i=1}^{k} N_i}}{\frac{\omega_i}{\sum_{i=1}^{k} \omega_i}} \tag{2}$$

Table 1 – Representative behaviors of FISA2 simulator.

State	PRZ indicators			S/G (A) indicators			S/G (B) indicators			Others
	L	P	T	L	FF	SF	L	FF	SF	
Normal										
LOCA	↓	↓	↓							
SGTR (A)	↓	↓	↓	↑	↓↓	↑	↓	↑	↓	
SGTR (B)	↓	↓	↓	↓	↑	↓	↑	↓↓	↑	
SLB (A)				↑	↓↓	↑↑	↑	↓	↑	
SLB (B)				↑	↓	↑	↑	↓↓	↑↑	
FLB (A)				↓	↑↑	↓	↑	↓	↑	
FLB (B)				↑	↓	↑	↓	↑↑	↓	

FF, feed flow; FLB (A), feed line break in loop A; FLB (B), feed line break in loop B; L, level; LOCA, loss of coolant accident; P, pressure; PRZ, pressurizer; SF, steam flow; S/G (A), steam generator in loop A; S/G (B), steam generator in loop B; SGTR (A), steam generator tube rupture in loop A; SGTR (B), steam generator tube rupture in loop B; SLB (A), steam line break in loop A; SLB (B), steam line break in loop B; T, temperature; ↑, increase; ↑↑, rapid increase; ↓, decrease; ↓↓, rapid decrease.

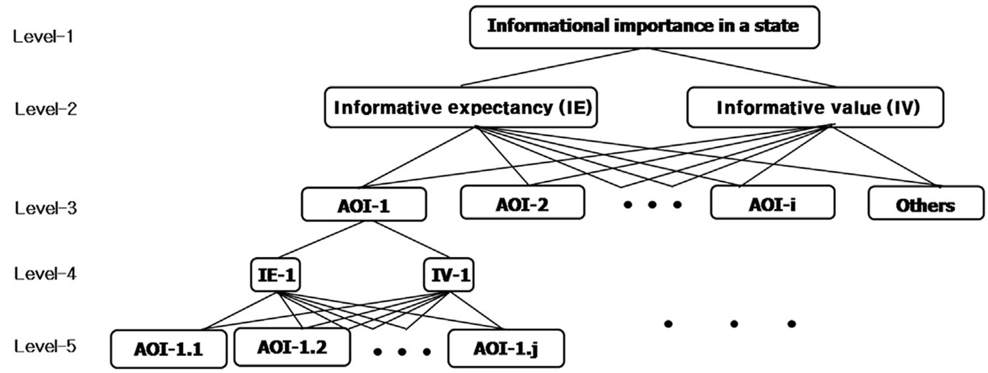

Fig. 1 – Setting up of the AHP for the evaluation of informational importance. AHP, analytical hierarchy process; AOI, area of interest; IE, informative expectancy; IV, informative value.

$$FIR^D(i) = \frac{\frac{D_i}{\sum_{i=1}^{k} D_i}}{\frac{\omega_i}{\sum_{i=1}^{k} \omega_i}} \tag{3}$$

$$FIR(i) = \frac{FIR^N(i) + FIR^D(i)}{2} \tag{4}$$

where,

$FIRN(i)$ = FIR with respect to number of fixations
$FIRD(i)$ = FIR with respect to duration of fixations
N_i = the number of eye fixation on information source-i
D_i = the duration of eye fixation on information source-i
k = total number of information sources
ω_i = importance of information source-i

A lot of studies on information searching or visual sampling behaviors have employed fixation frequency or/and fixation dwell time [10,13–16,26,40]. Fixation frequency, dwell time, or an integrated measure such as the average value of both could be used as a visual resource. In this study, the average value is employed for the calculation of FIR using Eq. (4). The relative attentional resources consumed on an information source should be equal to the relative importance of the information source in order to maximize the effectiveness of the attentional resources. Consequently, all FIR(i) should approach unity for the best effectiveness. The FIR is the ratio of relative attentional resources consumed on an information source to the relative importance of the information source. The FIR represents the attentional-resource effectiveness for each information source. The SAE incorporates the FIRs for all information sources, as follows:

$$SAE = \frac{\sum_{i=1}^{k} |FIR(i) - 1|}{k} \tag{5}$$

Hence, the SAE represents the overall attentional-resource effectiveness based on all information sources. The SAE should approach zero to maximize the overall attentional-resource effectiveness, because all FIR(i) should approach unity for the best effectiveness. An eye tracking system is used to obtain the eye fixation data (e.g., number and duration). The analytical hierarchy process (AHP) is applied to quantify the importance of information sources based on system behaviors [45].

2.3. *Quantification of informational importance*

Generally, normal operating conditions in NPPs are considered as safe conditions. No drastic change is observed during the normal operating conditions. Attention is paid to operators' monitoring behaviors during abnormal states such as accidents, incidents, or transient conditions. If an accident occurs in an NPP, operators should collect and integrate a symptom set of important information sources to correctly diagnose the current accident. The informational importance of an information source should be a function of its ability to differentiate between competing hypotheses (competing accidents) of the cause of a plant symptom. Hence, a set of the informational importance for an abnormal state is evaluated by considering a symptom set that has the ability to be diagnostic across a set of competing hypotheses. The following quotation from the report of Stubler et al [46] provides the basis for the quantification of the informational importance:

> "NPP operators learn through experience where and when to look in their work environment to gain the greatest information, and selectively focus attention on these sources. In dynamic environments, such as NPPs, there is a tendency for operators to attend to those sources that change most frequently (i.e., contain the most information in terms of bits per unit of time), or are likely to change given the current situation. These are examples of top-down processing (e.g., based on their understanding of the current situation, operators develop expectations of information sources that will provide the most useful information)."

Emphasis has been placed on the importance of information sources that change most frequently and are likely to change given a situation. In addition, it should be noted that information sources that do not change but have the ability to differentiate between competing hypotheses, such as the containment radiation in the preceding SGTR example, should be considered as important symptoms [47]. Hence, the attribute corresponding to the frequent change of the source (bandwidth) is denoted by an "informative value (IV)," and the attribute of the diagnostic symptom set including changing and unchanging symptoms is denoted by "informative expectancy (IE)."

Sets of the informational importance were evaluated for various accidents using the AHP, which generally involves the following four steps [45]:

- Step 1: Setting up the hierarchy by breaking the problem into a hierarchy of interrelated elements
- Step 2: Collecting input data by pairwise comparison of elements
- Step 3: Using the eigenvalue method to estimate the relative weights of elements
- Step 4: Aggregating the relative weights of elements to arrive at a set of relative importance for information sources

As shown in Fig. 1, a hierarchy is built for evaluating the informational importance. The informational importance of an accident (e.g., SGTR or LOCA) is placed at the top level, and divided into IE and IV at Level 2. An area of interest (AOI) refers to a region of HMIs that has information sources of importance. At Level 3, AOIs of important components [e.g., PRZ or steam generator (S/G)] are located. "Others" at Level 3 include all AOIs that do not have information sources important to the relevant accident. An AOI at Level 3 might break down into sub-AOIs such as indicator level, if the AOI at Level 3 has several indicators (e.g., level, pressure, or temperature). Development of the AHP depends on the quality of eye tracking measurements. If AOIs can be discriminated at an indicator level with an eye tracking system, a hierarchy with a depth of five levels can be developed, and another set of IE and IV are located at Level 4. If AOIs can be discriminated not at an indicator level but only at a component level, the hierarchy should have a depth of three levels.

After completing the hierarchy, input data (rating scales) are obtained to quantify the informational importance of AOIs. Input data comprise judgment matrices of pairwise comparisons in the AHP. Elements in one level are compared with each other in a pairwise manner according to their contribution to achieving the criteria of the next higher level. Input data of a judgment matrix are collected by questioning which one is more important or contributing to an element in the next higher level relatively to other one in the same level. If the ij^{th} element of the judgment matrix is inputted, the ji^{th} element is inputted by its reciprocal value. Since it is reciprocal and the diagonal elements are equal to unity, the number of pairwise comparisons required for a judgment matrix of order n is $n(n-1)/2$. A reliable scale should be developed to transform these qualitative statements into numbers for the quantification. The AHP provides a reliable scale, the definition of which is explained in Table 2.

The judgment matrix is used for evaluation of the relative importance of the elements in the AHP. The eigenvalue method proposed by Saaty [45] has been preferred for the

Table 2 – Numerical values and definition in the AHP scale [43].

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between adjacent scale values	When compromise is needed
Reciprocals of the preceding nonzero numbers	If activity i has one of the preceding nonzero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	A reasonable assumption
Rationales	Ratios arising from the scale	If consistency were to be forced by obtaining n numerical values to span the matrix

AHP, analytical hierarchy process.

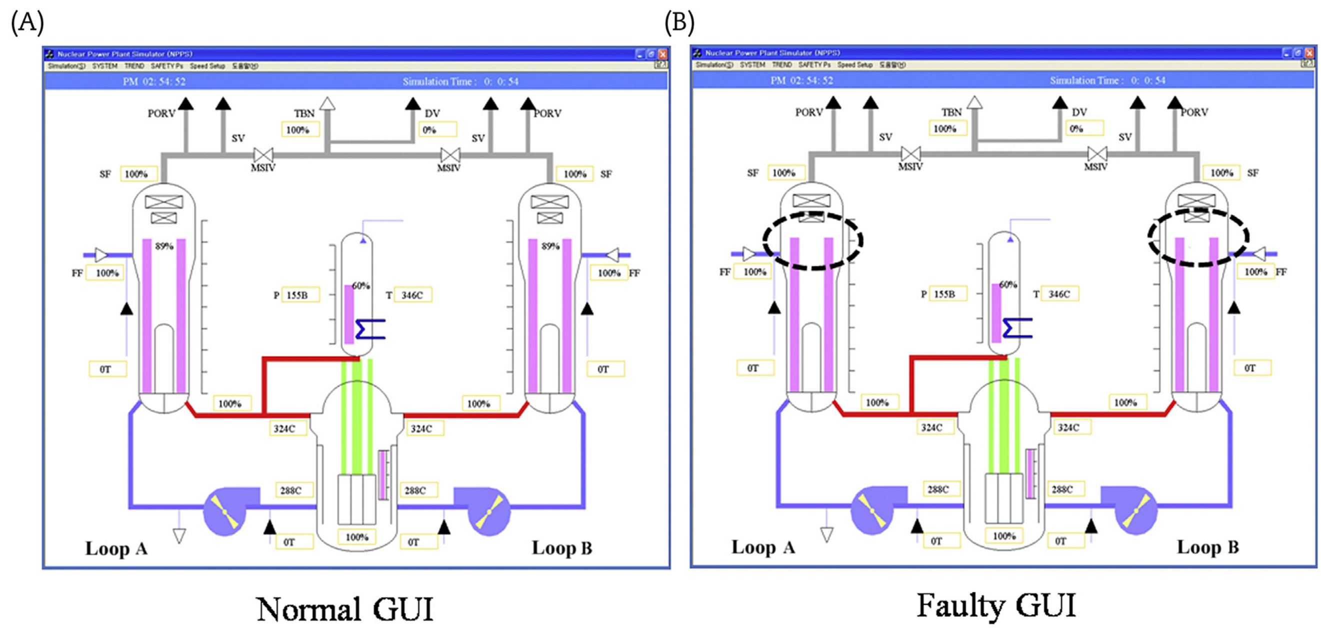

Fig. 2 – (A) Normal GUIs of FISA2 simulator. (B) Faulty GUIs of FISA2 simulator. DV, dump valve; FF, feed flow; GUI, graphic user interface; MSIV, main steam isolation valve; P, pressure; PORV, power operated relief valve; SF, steam flow; SV, safety valve; T, temperature; TBN, turbine.

evaluation. The informational importance of AOIs located at the bottom level is calculated by multiplying the relative weights from the top to the bottom elements.

In this study, values of the rating scale that were inputted into a judgment matrix were evaluated based on the behaviors of the system (i.e., an NPP system in this study). Hence, more reliable quantification is made compared with other AHP applications based only on the decision maker's expertise or opinion. A detailed process for the quantification used in this study is explained in the "Experiments" section.

2.4. *Inference of operators' thoughts from eye movement data*

Hume's epistemology provides a philosophical basis for the eye–mind hypothesis [48]. This hypothesis suggests that cognitive processing automatically takes place whenever information is visually perceived. It was postulated that eye fixations have a close correlation with visual attention [49]. Empirical evidence shows that eye movements are linked to attention as well [50,51]. The eye–mind hypothesis contrasts with the information-processing theory, suggesting that if information is perceived, a cognitive process is developed using some of the information, while the rest of the information is not included [52]. However, the eye–mind hypothesis is applicable to the context of complex industrial systems such as NPPs, which have representative system behaviors. Operators in such systems generally have well-developed knowledge (or mental models) of system behaviors through training and experience, and any information given to them is of potential significance for the operation of the system [53]. Usually, operators in NPP control rooms are highly experienced and trained periodically to establish the knowledge of system behaviors. The importance of individual information sources can be evaluated on the basis of the knowledge of system behaviors. There are thousands of information sources in NPPs, and even highly experienced operators cannot attend to all the information sources in abnormal situations. They pay selective attention to important information sources on the basis of their knowledge (mental model) to effectively understand the situation, which facilitates correct diagnosis of the situation. Effective selective attention to important symptoms or information sources should correspond with correct understanding and eventually correct diagnosis during abnormal situations, given that the operator have well-developed knowledge of system behaviors. This type of thought of an operator (understanding or diagnosis of a current situation) can be inferred from his or her selective attention pattern, which is analyzed based on the SAE evaluation. The better an operator selectively attends to information sources according to their informational importance, the lower the SAE value is expected to be, because the SAE should approach zero to maximize the overall attentional-resource effectiveness. If a diagnostic task with a specific accident simulation is given to an operator who has well-established knowledge of system behaviors, the operator is expected to make a good selective attention according to the importance of information sources for the specific accident. However, if the SAE evaluation is made using sets of informational importance of events other than the relevant accident, the SAE value would be higher than that with the sets of informational importance of the relevant accident. For example, if an SGTR occurs, an operator who has well-established knowledge of system behaviors should know which information sources are more to the SGTR and which are less important. The operator should pay selective attention to information sources that are important to the SGTR, to correctly understand and diagnose the situation. The operator's SAE calculated with the sets of informational importance for the SGTR should show a lower value (better selective attention) than the SAEs calculated with those for other accidents.

In this study, an accident is simulated with an NPP simulator, and each of the participating operators is required to monitor system behaviors, understand the situation, and finally report their diagnosis results. Eye movement data obtained during the simulation are used for the calculation of SAE with sets of informational importance for all possible accidents. The accident that is associated with the best SAE value (the lowest value) among all possible accidents should agree with the accident inserted into the simulation, given that the participant understands and diagnoses the situation correctly. The accident with the best SAE value (the lowest value) among all possible accidents is compared with that inserted into the simulation in each run of the simulation. The concordance rate (CR), which represents the agreement between the accident inserted into the simulation and that inferred from the best SAE evaluation, is calculated as a performance index of the inference of an operator's thought. Hence, the operator's thoughts (understanding or diagnosis of a current situation) can be inferred from his or her eye movement data by identifying the accident that shows the best SAE value (the lowest value) among all possible accidents.

3. Experiments

3.1. Pilot experiments

A pilot experiment was conducted to adjust the experiment setting with the FISA2 real-time simulator, which simulates a pressurized water reactor-type NPP [54]. An eye tracking system, FaceLAB 3.0 [55], was used to obtain eye fixation data such as eye fixation points and durations. The graphic user interface (GUI) of the FISA2 simulator was modified to enhance the measurement quality of the eye fixation data. Components and indicators placed too closely were relocated to be apart from each other, as shown in Fig. 2A. In addition, the FIR analysis revealed a problem in the existing GUI design of the FISA2 simulator. For all participants, regardless of their expertise level, the FIR of the S/G level indicators showed remarkably low values both in loop A and in loop B, which means that few eye fixations were made on S/G level indicators compared to their informational importance. The S/G levels were presented by a bar-type graph without any numerical value. Participants of the pilot experiment reported that since no numerical value had been provided on these indicators, it was very difficult to perceive the trend of level change. Hence, numerical indicators presenting the changed amount for the S/G level were added at the top of the bar indicator, as shown in Figs. 2A and 2B.

3.2. Objective, tasks, participants, and apparatus

This experimental study aims to see the feasibility of inference of operators' thoughts (understanding or diagnosis of a current situation) by analyzing the operators' eye movement patterns based on SAE evaluation. In this study, NPP accidents were simulated with the FISA2 simulator and participating operators were required to monitor system behaviors, understand the situation, and finally report their diagnosis results. Eye movement data were obtained with an eye tracking system, FaceLAB 3.0, during the simulation. The eye movement data were used to calculate the SAE values with sets of informational importance of all possible accidents, among which the accident coupled with the best SAE value (the lowest value) should correspond to the accident inserted into the simulation, given that the operator understands and diagnoses the situation correctly. CRs between the inserted accident and the inference results based on the best SAE evaluation were calculated for 15 operators according to their knowledge (mental model) levels. Two CRs, CR-1 and CR-2, were calculated in this study. CR-1 is a CR between the inserted accident and the inference results based on the best SAE evaluation. CR-2 is another CR-1 given that the operator has diagnosed the situation correctly. Hence, the CRs can be regarded as performance indexes on how well the proposed approach can infer the operators' thoughts (understanding or diagnosis of a current situation). The SAE values have meanings only when operators have well-developed knowledge of system behaviors. If an operator does not have enough knowledge of system behaviors, he or she may look at information sources without any intention, which is against the eye–mind hypothesis. To investigate this knowledge effect, the operator knowledge levels were controlled by a training program and the time interval (6 months) between experiment trials (i.e., Experiment 1 and Experiment 2). The HMI design effect, which was explained in the pilot experiment section, was also considered. The HMI design was controlled by the normal and the faulty HMI designs, as shown in Figs. 2A and 2B.

The plant system states were assumed to include the normal operating condition and the following seven accident conditions:

- LOCA: loss of coolant accident
- SGTR (A): steam generator tube rupture (loop A)
- SGTR (B): steam generator tube rupture (loop B)
- SLB (A): steam line break (loop A)
- SLB (B): steam line break (loop B)
- FLB (A): feed line break (loop A)
- FLB (B): feed line break (loop B)

Six tasks inclusive of SGTR (A) and SLB (B) were given to the study participants randomly. The SGTR (A) and SLB (B) cases were analyzed for the CR calculations because they were reported in the pilot experiments to be most difficult to diagnose among the tasks.

The participants were graduate students (14 males and 1 female) with 5.2 years of nuclear engineering background on average. They performed the role of operators in the experiments. They had normal or corrected-to-normal vision. The purpose, experimental procedure, and tasks were explained to them prior to the experiments. Diagnostic tasks were performed by them before and after the training in Experiments 1 and 2. All 15 participating operators from Experiment 1 returned for Experiment 2. After each experiment, a small

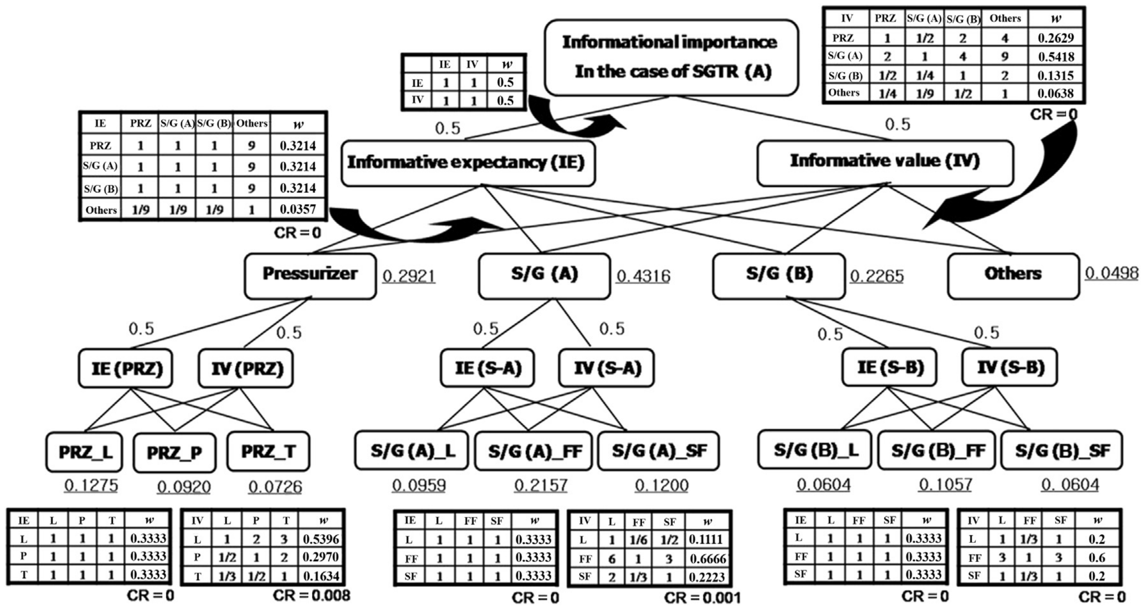

Fig. 3 – Setting up of the AHP for the SGTR (A). AHP, analytical hierarchy process; CR, concordance rate; FF, feed flow; IE, informative expectancy; IV, informative value; L, level; P, pressure; PRZ, pressurizer; S-A, steam generator A; S-B, steam generator B; SF, steam flow; S/G (A), steam generator in loop A; S/G (B), steam generator in loop B; SGTR (A), steam generator tube rupture in loop A; T, temperature.

interview was conducted to evaluate the operators' diagnosis results and their visual attention strategy.

Experiment 1 was conducted with the FISA2 simulator with the normal GUI, as shown in Fig. 2A. At the beginning of this study, only the knowledge effect was considered. However, after Experiment 1 the idea to consider the HMI design effect (normal vs. faulty GUI designs) came up and Experiment 2 was conducted with the normal and the faulty designs. The normal GUI design was identical to that used in Experiment 1. The faulty GUI design had the design deficiency found in the pilot experiment, as shown in Fig. 2B. In Experiment 2, the normal and the faulty GUI designs were applied to SGTR (A) and SLB (B), respectively. FaceLAB 3.0 was employed to measure the number and duration of eye fixations. As a lower bound of dwell times for the eye movement measurement, a figure of around 0.5 seconds was suggested for real-life tasks, although shorter times might be observed in some laboratory experiments when static, rather than dynamic, patterns are used as displays [33]. In scene perception and reading tasks, fixation duration ranged from 0.15 seconds to 0.6 seconds [56]. The fixation dwell time was 0.2–0.5 seconds in simple target tracing or detecting tasks [57–59]. In this study, AOIs were located at predefined areas, and only values in the AOIs were changed dynamically. Consequently, a value of lower than 0.5 seconds (i.e., the lower bound reported) was deemed to be more appropriate in defining the fixation. From the investigation of eye fixation data in the pilot test, a value of 0.25 seconds was empirically selected as a lower bound to operationally define a single fixation. A circular fixation area was employed in this study to define fixation. The fixation circle was larger than any indicator of the FISA2 simulator. If the fixation circle overlapped an indicator or an AOI for 0.25 seconds or more, it was counted as a single fixation.

3.3. *Quantification of informational importance*

The AHP was applied to quantify the relative importance of each information source. Behaviors of the system of interest are mainly described by indicators of process parameters and facilitate pairwise comparison between indicators for quantification. Table 1 provides a summary of behaviors of the simulator used in this study. One or more indicator(s) might be associated with a component. Component level is a higher level than the indicator level. Fig. 3 shows an example of the SGTR (A) case for the AHP. The overall hierarchy consists of five levels. The first level (Level 1) of informational importance has two second-level (Level 2) elements: IE and IV. There might be a controversy on whether the IE or the IV is more important to the informational importance. The bandwidth (change rate) will attract selective attention given that the IE (symptom sets) is not considered, which is a data-driven monitoring process [48]. On the other hand, the symptom sets (i.e., the situation–event relationships) will engage selective attention given that the IV (bandwidth) is ignored, which is considered to be a knowledge-driven monitoring process [38]. Since close correlations between process parameters are generally observed in NPPs, knowledge-driven monitoring should be considered as important as data-driven monitoring. Hence, even weighting 0.5 to both the IE and the IV is thought to be reasonable in NPPs. Behaviors of the FISA2 simulator were thoroughly analyzed to quantify the relative weights of components or indicators with respect to the IE and IV, respectively. The third level (Level 3) corresponds to the component level, whereas the fifth level (Level 5) represents the indicator level. The fourth level (Level 4) plays the same role as the second level and even weighting of 0.5 is given to both the IE (k) and the IV (k), where k = PRZ, steam generator A, or steam generator B. If the indicators of the pressurizer level (PRZ_L), pressure (PRZ_P), and temperature (PRZ_T) show decreases, three hypothetical accidents, LOCA, SGTR (A), or SGTR (B), are expected to compete. In order for operators to correctly diagnose the situation among the three competing hypotheses, additional information from the S/G (A) and S/G (B) indicators of L, feed flow (FF), and steam flow (SF)should be obtained as shown in Table 1. However, "Others" indicators in Table 1 that are not related to the PRZ, S/G (A), and S/G (B) do not change when PRZ_L, PRZ_P, and PRZ_T decrease. Sets of information provided by nine indicators of process parameters for the PRZ, S/G (A), and S/G (B) constitute the symptoms used for the diagnosis in this study. Hence, the nine indicators have the same importance in terms of the IE (see the 3 judgment matrices having all the elements of unity located in each left-side image at the bottom of Fig. 3). The three components PRZ, S/G (A), and S/G (B) have the same importance in terms of the IE between Levels 2 and 3, which leads to an input value of unity to the judgment matrix located in the upper left-hand side of Fig. 3. However, the "Others" in Fig. 3 do not include any significant symptoms, whereas each of PRZ, S/G (A), and S/G (B) does. Each of PRZ, S/G (A), and S/G (B) is much more important than the "Others," giving an input value of nine to the judgment matrix located in the upper left-hand side of Fig. 3. In terms of the IV, weights of components or indicators can be evaluated more easily. Observation of the system behaviors provides objective information on the bandwidth (change rate) of indicators. For instance, the change rate of PRZ_L was two times faster than that of PRZ_P and three times faster than that of PRZ_T, as shown in the second matrix from the left at the bottom of Fig. 3. All other

Table 3 – Sets of informational importance.

Components (AOIs)	LOCA	SGTR (A)	SLB (A)	FLB (A)
Informational importance at component level				
PRZ	0.5357	0.2921	0.1964	0.2122
S/G (A)	0.2024	0.4316	0.4822	0.4458
S/G (B)	0.2024	0.2265	0.2679	0.291
Others	0.0595	0.0498	0.0536	0.051
Informational importance at indicator level				
PRZ-L	0.2338	0.1275	0.0655	0.0707
PRZ-P	0.1688	0.092	0.0655	0.0707
PRZ-T	0.1331	0.0726	0.0655	0.0707
S/G (A)-L	0.0675	0.0959	0.1148	0.0907
S/G (A)-FF	0.0675	0.2157	0.2181	0.2267
S/G (A)-SF	0.0675	0.12	0.1492	0.1247
S/G (B)-L	0.0675	0.0604	0.0581	0.0491
S/G (B)-FF	0.0675	0.1057	0.125	0.1027
S/G (B)-SF	0.0675	0.0604	0.0848	0.1027
Others	0.0595	0.0498	0.0536	0.051

AOI, area of interest; FF, feed flow; FLB (A), feed line break in loop A; L, level; LOCA, loss of coolant accident; P, pressure; PRZ, pressurizer; SF, steam flow; S/G (A), steam generator in loop A; S/G (B), steam generator in loop B; SGTR (A), steam generator tube rupture in loop A; SLB (A), steam line break in loop A; T, temperature.

Table 4 – The t test results of the MMS.

Experiment (averaged MMS)	The t test results: t value, degree of freedom, p
1B (31.98) versus 2B (43.95)	$t = 3.25$, df = 14, $p < 0.01$
1B (31.98) versus 1A (88.37)	$t = 13.28$, df = 14, $p < 0.01$
2B (43.95) versus 1A (88.37)	$t = 12.42$, df = 14, $p < 0.01$
2B (43.95) versus 2A (97.26)	$t = 16.56$, df = 14, $p < 0.01$

df, degree of freedom; MMS, mental model score; 1A, Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; 2B, Experiment 2 before training.

weights to the IV can be evaluated in a similar manner. The set of the informational importance was then calculated by incorporating all the relative weights obtained from the judgment matrices. The same method was applied to calculate the sets of informational importance for LOCA, SLB (A), and FLB (A). Table 3 shows the sets of informational importance for LOCA, SGTR (A), SLB (A), and FLB (A). The sets of informational importance for SGTR (B), SLB (B), and FLB (B) are easily obtained by exchanging the informational importance of the relevant indicators of S/G (A) with those of S/G (B).

3.4. *Training and knowledge (mental model) evaluation*

The main experimental study included Experiments 1 and 2. The extent of an operator's knowledge was evaluated to see its effect. The operator's knowledge develops through a training program that the operator has undergone. Hence, the level of the operator's knowledge before training in Experiment 1 is considered to be poor. The operator's knowledge level after the training in Experiment 1 is deemed to be well constructed. In order to consider an intermediate level of operator knowledge (i.e., forgetting factor), Experiment 2 was conducted 6 months after Experiment 1. Knowledge fades with time as long as no training is provided. The level of operator's knowledge before the training in Experiment 2 is considered better than that before the training in Experiment 1 and poorer than that after the training in Experiments 1 and 2. The operator's knowledge before the training in Experiment 2 is, therefore, deemed as intermediately constructed. The training includes learning the behaviors of the FISA2 simulator and exercising the simulations by the participants, and testing of their knowledge. First, behaviors of the FISA2 simulator were presented and explained in detail to the participants. Second, they were acquainted with system behaviors with hands-on exercises. It should be noted that the participating operators were not trained to pay selective attention in a predefined way, but to be acquainted with system behaviors so that they could control their attention according to their knowledge. Finally, the operators required to explain system behaviors in each accident case, which was an evaluation test of their knowledge. This test was performed with questionnaires in order to see the completeness of the training. The operators passed the test only when they answered all the questions correctly. The training was iterated until a satisfactory test result was obtained. The effects of the training and time interval on the knowledge were investigated with four questionnaires on the system behaviors of

LOCA, SGTR (B), SLB (A), and FLB (B). Each questionnaire consisted of 20 questions. The participating operators were requested to answer how process parameters could be developed in a given accident scenario. The average score of the four questionnaires was used as a mental model score (MMS). The MMS was calculated before and after the training in Experiments 1 and 2.

3.5. *Experiment results*

Knowledge of an operator improves after training and worsens after a time interval during which the operator has no training or experience. The degree of the mental model was controlled as poor, intermediate, and well constructed by the training and the time interval (6 months). The MMS before the training was lower than that after the training in Experiments 1 and 2, as summarized in Table 4. The *t* tests based on pairwise comparison before and after the training were performed to statistically analyze the results of the MMSs. The *t* test results show that there is a statistically significant difference between the MMSs. The averaged MMSs improved in order of 1B (31.98), 2B (43.95), 1A (88.37), and 2A (97.26) (Table 4) (1A representing Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; and 2B, Experiment 2 before training).

CRs between the inserted accident and the inference results based on the best (i.e., lowest) SAE evaluation are summarized in Tables 5 and 6 for the SGTR (A) and SLB (B) cases, respectively. The CRs were calculated from the eye movement data at the component and indicator levels. AOIs at the component level include the PRZ, S/G (A), S/G (B), and other cases. AOIs at the indicator level include all the indicators provided.

In the SGTR (A) case, as the operator's knowledge of system behaviors improved in order of 1B (31.98), 2B (43.95), 1A (88.37), and 2A (97.26), the number of correct answers (diagnosis) increased, as shown in the CAR (correct answer rate) column in Table 5. In addition, CR-1, the CR between the inserted accident and the inference results based on the best SAE evaluation, became better. CR-2, the CR (CR-1) given that the operator has diagnosed the situation correctly, shows better results than CR-1 even when the operators had an intermediate level (2B) of knowledge about system behaviors. This means that even though the operators did not have enough knowledge of system behaviors, if they understand the situation correctly, a higher-accuracy inference can be made. In both experiments after the training (1A and 2A) no difference was observed between CR-1 and CR-2, because all the operators reported correct answers after the experiments. The CRs at the indicator level were evaluated better than those at the component level, which means that if more detailed information on operators' eye movement is obtained, a higher-accuracy inference can be made. In the SGTR (A) experiments, operators' thoughts (understanding or diagnosis of a current situation) can be inferred from their eye movement data with an accuracy of 80% at the component level and 93.33% at the indicator level, given that the operators have well-constructed knowledge such as in the 2A case.

Table 5 – Concordance rate for the SGTR (A) case.

Experiment	CAR (%)	CR-1 (%)	CR-2 (%)
Component level			
1B	13	0	0
2B	40	13.33	50.00
1A	100	73.33	73.33
2A	100	80.00	80.00
Indicator level			
1B	13	6.67	0
2B	40	53.33	83.33
1A	100	86.67	86.67
2A	100	93.33	93.33

CAR, correct answer rate; CR-1, concordance rate between the inserted accident and the inferred result from the best SAE evaluation; CR-2, concordance rate (CR-1) given that the operator has diagnosed the situation correctly; SAE, selective attention effectiveness; SGTR (A), steam generator tube rupture in loop A; 1A, Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; 2B, Experiment 2 before training.

Table 6 – Concordance rate for the SLB (B) case.

Experiment	CAR (%)	CR-1 (%)	CR-2 (%)
Component level			
1B	0	40.00	Not available
2B	26.67	66.67	100
1A	93.33	73.33	78.57
2A	100	73.33	73.33
Indicator level			
1B	0	53.33	Not available
2B	26.67	46.67	75.00
1A	93.33	80.00	85.71
2A	100	60	60

CAR, correct answer rate; CR-1, concordance rate between the inserted accident and the inferred result from the best SAE evaluation; CR-2, concordance rate (CR-1) given that the operator has diagnosed the situation correctly; SAE, selective attention effectiveness; SLB (B), steam line break in loop B; 1A, Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; 2B, Experiment 2 before training.

In the SLB (B) case, similar results were observed except for Experiment 2 (2B and 2A), where the faulty HMI design was used to investigate the faulty design effect, as shown in Table 6. As the operators' knowledge improved, the CAR increased and CR-1 improved except for Experiment 2. The design faults in S/G (A) and S/G (B) level indicators were intentionally inserted in the GUI in the SLB (B) cases, as shown in Fig. 2. The averaged FIRs for the 15 operators, which are used to investigate the faulty design effect, are given in Table 7. The FIR is the ratio of attentional resources spent on an information source to the importance of the information source. The FIR represents attentional-resource effectiveness in terms of each information source. Consequently, all FIRs should approach unity for the best effectiveness. The faulty design effect could be observed only at the indicator level, as shown in the FIR values of S/G (A)-L and S/G (B)-L for the SLB (B) cases in Table 7. The FIR values of S/G (A) and S/G (B) levels with the faulty HMI design show remarkably poor performance (bold and underlined figures) that is far from unity (FIRs < 0.5). In the debriefing after the experiments, most participating operators

Table 7 – Averaged FIRs on information sources.

Indicators (AOIs)	SGTR (A)				SLB (B)			
	1B	2B	1A	2A	1B	2B	1A	2A
Component level								
PRZ	1.06	0.91	1.01	0.96	0.89	1.04	0.96	1.07
S/G (A)	0.51	0.67	0.91	0.89	0.57	0.72	0.82	0.82
S/G (B)	0.74	1.12	0.91	1.12	0.66	0.81	0.92	0.96
Indicator level								
PRZ-L	1.31	1.40	1.28	1.31	1.38	2.0	1.41	1.83
PRZ-P	0.95	0.67	0.80	0.87	0.56	0.61	0.79	0.73
PRZ-T	0.77	0.73	0.81	0.69	0.66	0.71	0.67	0.74
S/G (A)-L	0.65	0.77	1.00	0.96	0.86	0.30	0.77	0.40
S/G (A)-FF	0.43	0.64	0.88	0.79	0.56	0.90	0.70	0.86
S/G (A)-SF	0.54	0.61	0.90	0.97	0.53	0.69	0.98	0.98
S/G (B)-L	0.71	1.05	0.82	1.19	0.71	0.39	0.89	0.45
S/G (B)-FF	0.72	1.08	0.87	0.84	0.67	1.04	0.85	1.03
S/G (B)-SF	0.75	1.14	1.07	1.36	0.65	0.90	1.04	1.29

AOI, area of interest; FF, feed flow; FIR, fixation to importance ratio; L, level; P, pressure; PRZ, pressurizer; SF, steam flow; S/G (A), steam generator in loop A; S/G (B), steam generator in loop B; SGTR (A), steam generator tube rupture in loop A; SLB (B), steam line break in loop B; 1A, Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; 2B, Experiment 2 before training.

reported that it was hard to perceive and maintain the change in S/G (A) and S/G (B) levels with the faulty HMI design. The operators adopted a strategy of focusing on the other two indicators, S/G FF and S/G SF, instead of focusing on the S/G level, because they thought that focusing on S/G FF and S/G SF would be more effective. This faulty design was made at the indicator level, and no remarkably poor performance (FIRs < 0.5) with respect to FIR values was observed at the component level, which leads to no remarkable decrease in the CR values in Experiment 2 compared with those in Experiment 1 at the component level, as shown in Table 6. However, at the indicator level, remarkable decreases in the CR values in Experiment 2 were observed. The CR-1 value was decreased from 53.33% in the 1B case to 46.67% in the 2B case, and from 80% in the 1A case to 60% in the 2A case, even though the operators' knowledge and CAR were improved. This means that if some design fault in the HMI exists, the inference from the best SAE evaluation might be compromised. In the SLB (B) experiments, operators' thoughts can be inferred from their eye movement data with an accuracy of 73.33% at the component level and 80% at the indicator level, given that the operators have well-constructed knowledge, such as the 1A case.

If only the experimental cases with well-constructed knowledge and no design fault in the HMI such as the SGTR (A) cases of 1A and 2A and the SLB (B) case of 1A are considered, operators' thoughts can be inferred from their eye movement data with an accuracy of 75.55% at the component level and 86.67% at the indicator level on average.

The t -tests were conducted to statistically analyze the difference between the mean of the SAE values of the 15 operators for the inserted accident and that for each of the other competing accidents, as shown in Table 8. The mean of the SAE values for the inserted accident should be lower than that for each of the other competing accidents. In the poor knowledge (mental model) cases of 1B at both the component and the indicator levels, no statistically significant differences were observed between the mean of the inserted accident and that of each of other competing accidents, except for the SLB (B) case at the indicator level. Operators' eye movement is likely to be governed by the bandwidth (change rate) when operators do not have sufficient knowledge of system behaviors. Behaviors of the indicators of LOCA, SGTR (A), and FLB (A) are considerably different from those of SLB (B) in terms of bandwidth, as shown in Table 1, which is thought to result in the statistical differences with α (significance level) $= 0.05$ between the mean of SLB (B) and that of LOCA, SGTR (A), and FLB (A). As the level of knowledge improves in order of 1B, 2B, 1A, and 2A, more cases of statistical differences between the means were observed. This implies that if operators have better knowledge, the inference based on the best SAE evaluation is made more precisely. More cases of statistical differences between the means were observed at the indicator level than at the component level, which supports the idea that if more detailed information on operators' eye movement is obtained, a higher-accuracy inference can be made. In the well-developed knowledge cases of 2A, especially at the SGTR (A) indicator level, all the instances show statistically significant differences between the means (i.e., 5 instances with $\alpha = 0.01$ and one instance with $\alpha = 0.05$), as shown in Table 8. Even in 1A case of the SGTR (A) indicator level, five instances out of six show a statistical difference between the means with $\alpha = 0.01$. However, at the component level in the SGTR (A) cases, the means between SGTR (A) and each of SLB (A) and FLB (A) do not show significant differences, because their behaviors at the component level are slightly similar in terms of bandwidth and expectancy, as can be inferred from Table 1. In the cases of SLB (B), the behaviors of indicators of SLB (B) also show a slightly similar pattern to those of SGTR (B) and FLB (B) in terms of bandwidth and expectancy. Hence, they do not show significant difference in their means of the SAE evaluation. Otherwise, statistically significant differences were observed in the SLB (B) cases such as 2B, 1A, and 2A (Table 8).

4. Discussion

4.1. *Applicable areas of the proposed inference method*

In this study, a novel method for inference of operators' thoughts (understanding or diagnosis of a current situation) from their eye movement data was proposed and evaluated with experiments using an NPP simulator. This method is expected to be effectively applied to enhancing the safety of NPP operations. In safety-critical and complex systems such as NPPs, human errors have been considered as a serious cause of accidents, especially after the Tree Mile Island accident. There have been two general approaches to cope with human errors in NPP control rooms. The first approach is to develop well-constructed training programs to which the inference method based on operators' eye movement data can be effectively applied. Operators' understanding or diagnosis during a simulation training can be monitored in real time with the inference method. Advice and/or recommendation

Table 8 – Statistical analysis (t test) of means of the SAE evaluation between inserted accident and other accidents.

Inserted accident	Other accident	Statistics (t value, p), degree of freedom = 28			
		1B	2B	1A	2A
Component level					
SGTR (A)	LOCA	(1.57, 0.06)	(0.40, 0.35)	(1.57, 0.06)	(6.25, 0.00)[a]
	SGTR(B)	(0.03, 0.49)	(0.49, 0.32)	(2.13, 0.02)[b]	(3.68, 0.00)[a]
	SLB (A)	(0.18, 0.43)	(0.08, 0.47)	(0.98, 0.17)	(1.51, 0.07)
	SLB (B)	(0.10, 0.46)	(0.34, 0.37)	(2.53, 0.01)[a]	(4.00, 0.00)[a]
	FLB (A)	(0.38, 0.35)	(0.18, 0.43)	(0.89, 0.19)	(1.18, 0.12)
	FLB (B)	(0.31, 0.38)	(0.33, 0.37)	(2.00, 0.03)[b]	(3.03, 0.00)[a]
SLB (B)	LOCA	(0.31, 0.38)	(0.97, 0.17)	(2.82, 0.00)[a]	(3.94, 0.00)[a]
	SGTR (A)	(1.47, 0.08)	(2.86, 0.00)[a]	(3.11, 0.00)[a]	(3.50, 0.00)[a]
	SGTR (B)	(0.51, 0.31)	(0.22, 0.41)	(0.47, 0.32)	(0.51, 0.31)
	SLB (A)	(0.52, 0.30)	(2.17, 0.02)[b]	(2.01, 0.03)[b]	(2.99, 0.00)[a]
	FLB (A)	(0.80, 0.22)	(2.08, 0.02)[b]	(1.86, 0.04)[b]	(2.55, 0.01)[a]
	FLB (B)	(0.46, 0.32)	(0.65, 0.26)	(0.36, 0.36)	(0.44, 0.33)
Indicator level					
SGTR (A)	LOCA	(0.35, 0.37)	(1.09, 0.14)	(3.92, 0.00)[a]	(6.72, 0.00)[a]
	SGTR (B)	(0.18, 0.43)	(1.02, 0.16)	(3.95, 0.00)[a]	(3.74, 0.00)[a]
	SLB (A)	(1.55, 0.07)	(2.12, 0.02)[b]	(0.98, 0.17)	(2.54, 0.00)[a]
	SLB (B)	(1.61, 0.06)	(2.86, 0.00)[a]	(6.75, 0.00)[a]	(4.64, 0.00)[a]
	FLB (A)	(1.55, 0.07)	(2.13, 0.02)[b]	(3.52, 0.00)[a]	(2.41, 0.01)[b]
	FLB (B)	(1.58, 0.06)	(3.14, 0.00)[a]	(6.89, 0.00)[a]	(4.85, 0.00)[a]
SLB (B)	LOCA	(1.91, 0.03)[b]	(3.64, 0.00)[a]	(6.35, 0.00)[a]	(5.87, 0.00)[a]
	SGTR (A)	(2.16, 0.02)[b]	(2.91, 0.00)[a]	(5.53, 0.00)[a]	(3.75, 0.00)[a]
	SGTR (B)	(0.32, 0.38)	(0.88, 0.19)	(0.89, 0.19)	(0.12, 0.45)
	SLB (A)	(1.69, 0.05)	(3.20, 0.00)[a]	(4.05, 0.00)[a]	(3.29, 0.00)[a]
	FLB (A)	(2.30, 0.02)[b]	(3.48, 0.00)[a]	(4.03, 0.00)[a]	(3.39, 0.00)[a]
	FLB (B)	(0.50, 0.31)	(0.46, 0.33)	(0.28, 0.39)	(0.58, 0.28)

[a] $p < 0.01$.
[b] $p < 0.05$.
FLB (A), feed line break in loop A; FLB (B), feed line break in loop B; LOCA, loss of coolant accident; SAE, selective attention effectiveness; SGTR (A), steam generator tube rupture in loop A; SGTR (B), steam generator tube rupture in loop B; SLB (A), steam line break in loop A; SLB (B), steam line break in loop B; 1A, Experiment 1 after training; 1B, Experiment 1 before training; 2A, Experiment 2 after training; 2B, Experiment 2 before training.

can be provided in a timely manner based on the inference results. In addition, operators' eye movement pattern can be evaluated and operators can be trained to give their best performance.

The second approach to reduce human errors in NPP control rooms is to design the NPP control room with improved interfaces and operator support systems. Operator support systems refer to the systems that provide useful information to operators or automated systems used for preventing human errors [3]. A new type of operator support system can be developed based on the inference method proposed in this study. An NPP is operated by a shift consisting of several operators including a supervisor. When an abnormal situation occurs, the supervisor needs to check other operators' understanding of the current situation. The inference method can be applied to supply real-time information on other operators' understanding of the current situation to the supervisor. This kind of information can help the supervisor make a diagnosis effectively when his or her diagnosis result agrees with other operators' results, or monitor human errors that might be committed by other operators with incorrect understanding. This kind of human error associated with incorrect understanding has been considered as one of the most significant human errors in NPPs.

In addition, this method can be applied to human factor validation, which is called integrated system validation in nuclear industries, for the assessment of operators' monitoring and diagnosis performance. The objective of the integrated system validation is to provide evidence that the integrated system adequately supports plant personnel in the safe operation of the relevant NPP [60]. Operators' tasks are generally completed with cognitive activities such as monitoring and detection of the environment, situation assessments, response planning, and response implementing [1]. Hence, the proposed method can be used for the assessment of operators' monitoring and situation assessment abilities in the integrated system validation.

4.2. *Fidelity and limitations of the study*

The main objective of this study is to show the feasibility of inference of an operator's thoughts (understanding or diagnosis of a current situation) by analyzing the operator's eye movement pattern using the SAE evaluation. As a first attempt, a low fidelity simulator was used. The FISA2 simulator was operated only on a single screen. In a full-scope simulator, which has a lot of alarms and displays, navigation would be made through various display screens as well as within a single screen to search for information sources important to a situation. In this experimental study, navigation among screens was not needed because a single-screen simulator was used. Obviously, navigation among screens is expected to be costlier than that within a single screen. Generally, information sources important to a situation were frequently fixated during the experiments. However, information sources not important to a situation were also fixated by chance at a greater rate than expected. Eye fixation on a single screen is relatively inexpensive. The operators who participated in this study sometimes fixated on unimportant information sources by chance, which is deemed as a type of noise detrimental to the experimental results. Navigation is thought to reduce this type of noise, because operators are likely to spend their time focusing on important information sources instead of looking at unimportant ones by chance. However, the FIR and the SAE might suffer from combinatorial explosion due to a huge number of information sources available for search in a full-scope simulator. Even though operators working in commercial NPPs are generally highly trained and experienced, and they have well-trained eye scanning patterns in abnormal situations, they might fixate on an unimportant information source by chance because of the large number of information sources presented in control rooms.

As an abnormal situation in NPPs is usually detected by the onset of alarms, the alarm design definitely affects eye movements. In this study, only indicators for process parameters are considered to see the feasibility of the method. Further experimental studies considering alarm designs are required. It should be noted that the participants of this experimental study are not well-trained operators even though they are graduate students majoring in nuclear engineering disciplines. Further experiments with field operators are required as well.

4.3. *Considerations for real-world applications*

In most human factor studies or training in NPPs, operational scenarios are determined in advance. Hence, operational situations are analyzed by evaluators (i.e., subject matter experts or human factor experts) in advance. Attention should be paid to the possibility that an operational strategy of operators might differ from the optimal strategy analyzed by evaluators. Sets of informational importance are obtained based on the optimal strategy. Hence, if operators adopt a different strategy, the informational importance should be modified taking into consideration the different strategy. Evaluators should carefully monitor operators' selective attention during scenarios and then analyze with operators after the scenarios. If a different strategy is adopted by operators, sets of informational importance should be modified accordingly. The FIR and SAE should be re-evaluated with the modified sets of informational importance. This kind of approach was applied in an evaluation of personnel task performance [61]. An optimal task solution was prepared in advance by evaluators. If deviations were identified during the simulation, these deviations were reflected in the modification of the optimal task solution.

5. Conclusion

In this experimental study, a novel method for inference of an operator's thoughts (understanding or diagnosis of a current situation) from his or her eye movement data is proposed and evaluated with an NPP simulator. The inference method was developed based on the SAE evaluation, which represents how effectively an operator attends to important information sources. CRs between the simulated accident and the inference results based on the SAE evaluation were calculated for 15 operators. The CRs can be regarded as performance indexes representing how well the proposed method can infer the operator's thoughts from his or her eye movement data. In the experiments, about 80% of the operator's thoughts can be correctly inferred using the proposed method. Hence, it is concluded that the inference method has a great potential for useful applications in NPPs, such as development of an improved operator training program, a new type of operator support system, and human performance measures for human factor validation.

Conflicts of interest

All authors have no conflicts of interest to declare.

Acknowledgments

This work was supported by the project of "Suitability Evaluation of Main Control Room in APR-1400 Nuclear Power Plant" under a grant from the Khalifa University Internal Research Fund (KUIRF; Fund #210021).

REFERENCES

[1] M. Barriere, D. Bley, S. Cooper, J. Forester, A. Kolaczkowski, W. Luckas, G. Parry, A. Ramey-smith, C. Thompson, D. Whitehead, J. Wreathall, Technical Basis and Implementation Guidelines for a Technique for Human Event Analysis (ATHEANA), Rev.01, NUREG-1624, US NRC, 2000.
[2] J. Kemeny, The Need for Change: the Legacy of TMI. Report of the President's Commission on the Accident at Three Miles Island, Pergamon, New York, 1979.
[3] H. Yoshikawa, Human–machine interaction in nuclear power plants, Nucl. Eng. Technol. 37 (2005) 151–158.
[4] S.J. Lee, P.H. Seong, Development of an integrated design support system to aid cognitive activities of operators, Nucl. Eng. Technol. 39 (2007) 703–716.
[5] T.B. Sheridan, Telerobotics, Automation, and Human Supervisory Control, MIT Press, Cambridge, MA, 1992.
[6] R.J. Mumaw, E.M. Roth, K.J. Vicente, C.M. Burns, there is more to monitoring a nuclear power plant than meets the eye, Hum. Factors 42 (2000) 36–55.
[7] A.L. Yarbus, Eye Movements and Vision, Plenum, New York, 1967 (originally published in Russian in 1962).
[8] M. Wang, Y. Maeda, Y. Takahashi, Human Intention Recognition via Eye Tracking Based on Fuzzy Inference, the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium

on Advanced Intelligence Systems (SCIS-ISIS 2012), Kobe, Japan, November 2012, pp. 846–851.
[9] Z. Huiping, I. Makoto, I. Toshiyuki, Eye movement-based inference of truck driver's intent of changing lanes, SICE J. Control Meas. Syst. Integr. 2 (2009) 291–298.
[10] J.S. Ha, P.H. Seong, Attentional-resource effectiveness measures in monitoring and detection tasks in NPPs, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 40 (2010) 993–1008.
[11] C.D. Wickens, J. Lee, Y. Liu, S.G. Becker, An Introduction to Human Factors Engineering, second ed., Prentice Hall, 2004.
[12] C.D. Wickens, J.G. Hollands, Engineering Psychology and Human Performance, third ed., Prentice Hall, Upper Saddle River, NJ, 2000.
[13] R.E. Jones, J.L. Milton, P.M. Fitts, Eye Movements of Aircraft Pilots: I. A Review of Prior Eye Movement Studies and a Description of Eye Movements during Instrument Flight, USAF Tech. Rep. No. 5837, Wright-Patterson Air Force Base, Dayton, OH, 1949.
[14] J.L. Milton, R.E. Jones, P.M. Fitts, Eye Movements of Aircraft Pilots: II. Frequency, Duration, and Sequence of Fixation when Flying the USAF Instrument Low Approach System (ILAS), USAF Tech. Rep. No. 5839, Wright-Patterson Air Force Base, Dayton, OH, 1949.
[15] P.M. Fitts, R.E. Jones, J.L. Milton, Eye movements of aircraft pilots during instrument landing approaches, Aeronaut. Eng. Rev. 9 (1950) 1–5.
[16] J.W. Senders, The human operator as a monitor and controller of multidegree of freedom system,, IEEE Trans. Hum. Factors Electron 5 (1964) 3–5.
[17] J.R. Carbonell, A queueing model for many-instrument visual sampling, IEEE Trans. Hum. Factors Electron HFE-7 (1966) 157–164.
[18] J.R. Carbonell, J.L. Ward, J.W. Senders, A queuing model of visual sampling: experimental validation, IEEE Trans. Man Mach. Syst. MMS-9 (1968) 82–87.
[19] R.D. Smallwood, Internal models of the human instrument monitor, IEEE Trans. Hum. Factors Electron HFE-8 (1967) 181–187.
[20] S. Baron, D.L. Kleinman, The human as an optimal controller and information processor, IEEE Trans. Man Mach. Syst. MMS-10 (1969) 9–17.
[21] S. Baron, D.L. Kleinman, W.H. Levison, An optimal control model of the human response: part ii: prediction of human performance in a complex task, Automatica 6 (1970) 371–383.
[22] D.L. Kleinman, R.E. Curry, Some new control theoretic models for human operator display modelling, IEEE Trans. Syst. Man Cybern. SMC-7 (1977) 778–784.
[23] D.L. Kleinman, S. Baron, W.H. Levison, An optimal control model of the human response: part i: theory and validation, Automatica 6 (1970) 357–369.
[24] W.H. Levison, A Control Theory Model for Human Decision Making, NTIS No. N73-10104, 7th NASA Annual Conference on Manual Control, University of Southern California, 1971.
[25] P. Wewerinke, A Model of the Human Decision Maker Observing a Dynamic System, Tech. Rep. NLR TR 81062 L, National Lucht-en Ruimtevaartlaboratorium, Netherlands, 1981.
[26] C.D. Wickens, J. Helleberg, J. Goh, X. Xu, W.J. Horrey, Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning, Tech. Rep. ARL-01-14/NASA-01-7, NASA Ames Research Center, Moffett Field, CA, 2001.
[27] T.B. Sheridan, On how often the supervisor should sample, IEEE Trans. Syst. Sci. Cybern. SSC-6 (1970) 140–145.
[28] T. Kvalseth, Quantitative Modeling of the Time-multiplexing Characteristics of Human Controllers, NSF GK-37419, 1975.
[29] T. Kvalseth, Human information processing in visual sampling, Ergonomics 21 (1977) 439–454.
[30] T. Kvalseth, The effect of cost on the sampling behavior of human instrument monitors, in: T.B. Sheridan, G. Johannsen (Eds.), Monitoring Behavior and Supervisory Control, Plenum, New York, 1977.
[31] N. Moray, Human Information Processing and Supervisory Control, Tech. Rep., Man-machine Systems Laboratory, MIT, Cambridge, MA, 1980.
[32] N. Moray, The role of attention in the detection of errors and the diagnosis of failure in man-machine systems, in: J. Rasmussen, W.B. Rouse (Eds.), Human Detection and Diagnosis of System Failures, Plenum, New York, 1981.
[33] N. Moray, Monitoring behavior and supervisory control, in: K. Boff, L. Kaufman, J. Thomas (Eds.), Handbook of Human Perception and Performance, John Wiley and Sons, New York, 1986.
[34] N. Moray, G. Neil, C. Brophy, The Behavior and Selection of Fighter Controllers, Tech. Rep., Ministry of Defense, London, 1983.
[35] E.R.F.W. Crossman, J.E. Cooke, R.J. Beishon, Visual attention and sampling of displayed information in process control, in: E. Edwards, F. Lees (Eds.), The Human Operator in Process Control, Taylor and Francis, London, 1974.
[36] J.W. Senders, Visual Scanning Process, University of Tilburg Press, Netherlands, 1983.
[37] A.X. Miao, G.L. Zacharias, S.P. Kao, A computational situation assessment model for nuclear power plant operations, IEEE Trans. Syst. Man Cybern. PART A 27 (1997) 728–742.
[38] M.C. Kim, P.H. Seong, A computational model for knowledge-driven monitoring of nuclear power plant operators based on information theory, Reliab. Eng. Syst. Saf. 91 (2006) 283–291.
[39] M.C. Kim, P.H. Seong, An analytic model for situation assessment of nuclear power plant operators based on Bayesian inference, Reliab. Eng. Syst. Saf. 91 (2006) 270–282.
[40] J.S. Ha, P.H. Seong, Experimental investigation between attentional-resource effectiveness and perception and diagnosis in nuclear power plants, Nucl. Eng. Des. 278 (2014) 758–772.
[41] D. Gopher, The skill of attentional control: acquisition and execution of attention strategies, in: D.E. Meyer, S. Kornblum (Eds.), Attention and Performance, Vol. 14, MIT Press, Cambridge, MA, 1993.
[42] D. Kahneman, Attention and Effort, Prentice Hall, Englewood Cliffs, NJ, 1973.
[43] J.M. O'Hara, J.C. Higgins, W.F. Kramer, J. Kramer, Computer-based procedure systems: technical basis and human factors review guidance, NUREG/CR-6634, US NRC, 2002.
[44] J. Rasmussen, Diagnostic reasoning in action, IEEE Trans. Syst. Man Cybern. 23 (1993) 981–992.
[45] T.L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, 1980.
[46] W.F. Stubler, J.M. O'Hara, J.C. Higgins, J. Kramer, Human Systems Interface and Plant Modernization Process: Technical Basis and Human Factors Review Guidance, NUREG/CR-6637, US NRC, 2000.
[47] J.S. Ha, P.H. Seong, A review of studies on operator's information searching behavior for human factors studies in NPP MCRs, Nucl. Eng. Technol. 41 (2009) 247–270.
[48] G. Underwood, J. Everatt, The role of eye movements in reading: some limitations of the eye–mind assumption, in: E. Chekaluk, K.R. Llewellyn (Eds.), The Role of Eye Movements in Perceptual Process, Advances in Psychology, Vol. 88, Elsevier Science Publishers BV, Amsterdam, 1992, pp. 111–169.
[49] N. Moray, I. Rotenberg, Fault management in process control: eye movements and action, Ergonomics 32 (1989) 1319–1342.

[50] T.L. Hodgson, H.J. Muller, Evidence relating to premotor theories of visuspatial attention, in: J.M. Findlay, et al. (Eds.), Eye Movement Research, Elsevier Science BV, 1995.
[51] W.X. Shneider, H. Deubel, Visual attention and saccadic eye movements: evidence for obligatory and selective spatial coupling, in: J.M. Findlay, et al. (Eds.), Eye Movement Research, Elsevier Science BV, 1995.
[52] J.B. Best, Cognitive Psychology, West Publishing Company, St. Paul, MN, 1992.
[53] A. Drøivoldsmo, J.G. Skraaning, M. Sverrbo, J. Dalen, T. Grimstad, G. Andresen, Continuous measure of situation awareness and workload, HWR-539, OECD Halden Reactor Project, 1988.
[54] H. Seok, H.C. No, Development of a software for the micro simulator for Kori Nuclear Power Plant Unit 2, Nucl. Technol. 106 (1994) 384–396.
[55] http://www.seeingmachines.com/.
[56] D.E. Irwin, Visual memory within and across fixations, in: K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading, Springer-Verag, New York, NY, 1992.
[57] J. Duncan, R. Ward, K. Shapiro, Direct measurement of attentional dwell time in human vision, Nature 369 (1994) 313–315.
[58] C.M. Moore, H. Egeth, L.R. Berglan, S.J. Luck, Are attentional dwell times inconsistent with serial visual search? Psychon, Psychon. Bull. Rev 3 (1996) 360–365.
[59] R. Ward, J. Duncan, K. Shapiro, Effects of similarity, difficulty, and non-target presentation on the time course of visual attention, Percept, Psychophys 59 (1997) 593–600.
[60] J.M. O'Hara, W.F. Stubler, J.C. Higgins, W.S. Brown, Integrated System Validation: Methodology and Review Criteria, NUREG/CR-6393, US NRC, 1997.
[61] G.J. Skraning, The Operator Performance Assessment System (OPAS), HWR-538, OECD Halden Reactor Project, 1998.