

# Effect of Speech Display on Team Mutual Awareness and Diagnosis Performance

Yingzhi Zhang<sup>(✉)</sup> and Zhizhong Li

Department of Industrial Engineering,  
Tsinghua University, Beijing, People's Republic of China  
zhangyz\_0208@163.com, zzli@tsinghua.edu.cn

**Abstract.** In many complex industrial systems, a team works together on digital systems. Team members need to gather information efficiently to maintain mutual awareness and do well in dealing with what may happen in the system. This research studies the effect of speech display in mutual awareness tools on team mutual awareness, individual performance, mental workload and team performance in simulated emergent diagnosis tasks of nuclear power plants at different task complexity levels. An experiment was conducted with 48 subjects (24 teams). Every team completed two scenarios. The results showed that speech display improved team mutual awareness in the low-complexity scenario. However, the teams without speech display did better in analyzing the accident than the teams with speech display in the low-complexity scenario. Furthermore, the teams with speech display were under higher workload.

**Keywords:** Speech display · Mutual awareness · Team performance · Complex industrial system

## 1 Introduction

In many complex industrial systems, team members work together to ensure the effective operation of the system. They have to respond quickly to any emergencies that may occur in the system. So besides having a real-time understanding of how their own part operates, they also need to keep an up-to-the-minute understanding of how other team members interact with the system, which is related to mutual awareness.

Situation awareness is defined as “the knowledge of a dynamic organization structure maintained through perceptual information gathered from the environment” [1]. When a team works together, one sort of situation awareness, called mutual awareness, is “A’s awareness of activities of B, including the aim, state, possible effects and ramifications etc.” [2]. For diagnosis tasks, including mutual awareness tools can improve mutual awareness and thus improve team performance to some degree [11].

To maintain mutual awareness and do well in dealing with what may happen in the system, team members need to gather information efficiently. The selection of information display is an important factor that influences the efficiency of information gathering.

Our research studied the effect of speech display in mutual awareness tools on team mutual awareness, individual performance, mental workload, and team diagnosis

performance at two levels of task complexity. The speech display was used only in mutual awareness tools and there is no voice prompt for one's own states.

Five questions are to be answered:

*Question 1:* How will speech display influence team mutual awareness?

*Question 2:* How will speech display influence individual performance?

*Question 3:* How will speech display influence mental workload?

*Question 4:* How will speech display influence team performance?

*Question 5:* Will the conclusion differ at different task complexity?

## 2 Literature Review

### 2.1 Visual, Auditory and Multimodal Display

In most complex systems, visual displays are most often used. However, there are some limitations in visual display modality and thus auditory display modality can serve as a complementary choice in addition to visual display modality to enhance performance in complex tasks and reduce cognitive load [3].

Also, researchers have stressed the use of multimodal display that combines visual display and auditory display.

The advantages of multimodal display include: synergy (presenting various dimensions of information of the same event), redundancy (presenting information in various ways), and the increase of bandwidth of information transfer [4].

Dowell et al. (2008) described three situations where multimodal display can be applied: multitask situations, usually with visual demand for primary tasks and auditory demand for side tasks, multimedia presentations, which present the information through multiple channels, and situations when the screen is too small to present all information visually [5].

### 2.2 Selection of Information Display Modality

Various experiments analyzed the influence of information display modality selection on individual performance, situation awareness and mental workload based on different scenarios.

One example is an experiment conducted by Kalyuga et al. (2009). Visual-only display, auditory-only display, and redundant (both visual and auditory) display were used to present information. Redundant display resulted in the best learning performance, followed by auditory display and visual display [6].

As Moreno and Mayer (2002) suggested, the reason why redundant display results in better performance may lie in the "sharing of load across visual and auditory processing in working memory" [7].

Dowell et al. (2008) conducted an experiment to examine the effect of information display on comprehension performance, taking information complexity into consideration. The results showed that redundant display and visual display did an equally good job and speech display results in worse performance in high information complexity [5].

Examples above are all single-task situations. In some occasions, a side task is performed while the primary task (usually an ongoing visual task) is processed, which is called the dual-task situation. There are two views on selection between “visual-visual” display (visual primary task and visual side task) and “visual-auditory” display (visual primary task and auditory side task) in dual-task situations.

Some argue that “visual-auditory” display can result in better performance. Wickens explained this using models of multiple resources. The resource management of different channels is relatively independent and thus an auditory side task interfere with the visual primary task less than a visual side task [8].

However, there exists an opposite view. As the primary visual task is ongoing and the auditory task is discrete, some assert that the auditory side task will take attention away from the visual primary task easily and adversely affect the performance of the primary task, which is called a “preemption” effect [9]. So they think “visual-visual” display is a better choice.

Horrey & Wickens (2004) took both views into consideration and posited that this is a trade-off which needs to be balanced when designing information display in dual-task situations [10].

### 3 Method

In our research, speech display was included in a simulated simplified nuclear power plant. An experiment was conducted with 48 participants (24 teams). Every pair of participants acted as two operators working together in the nuclear power plant. They were in charge of the nuclear island (Fig. 1) and the conventional island (Fig. 2), respectively. They were asked to complete some daily work alone and some diagnosis tasks together.

#### 3.1 Participants

The participants were 48 male undergraduate students major in science or engineering, with an average age of 21.2. Most of them knew little about nuclear power plants before participation. To avoid the influence of unfamiliarity, every pair of participants who know each other well were grouped into a team. Participants in one team all have known each other for more than 0.5 years and 62.50 % have known each other for 3–4 years.

The participants were randomly divided into two groups. One used system with speech display that would “speak out” the other team member’s actions and alarms. The other used system without speech display.

#### 3.2 System

Every pair of the participants conducted the experiment on two computers in a quiet room. The system was a simulated simplified nuclear power plant with a mutual

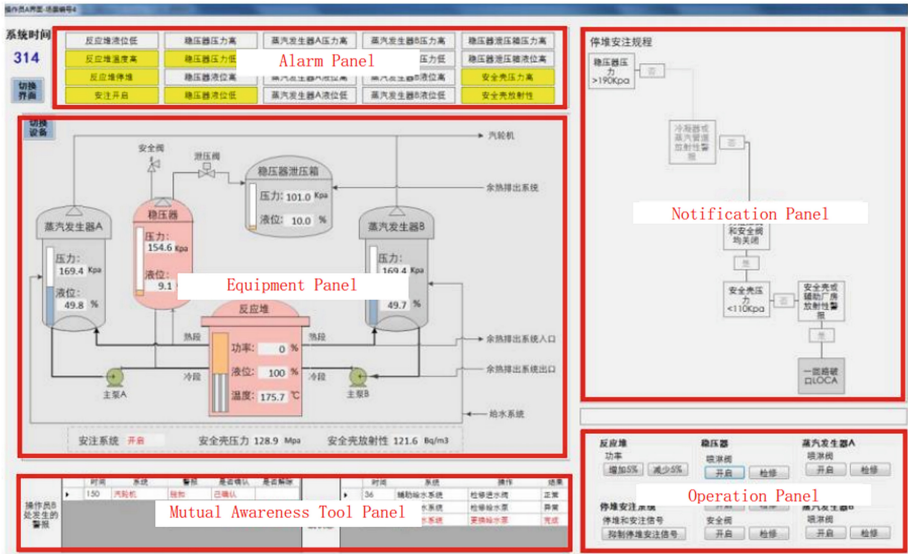


Fig. 1. Interface for nuclear island

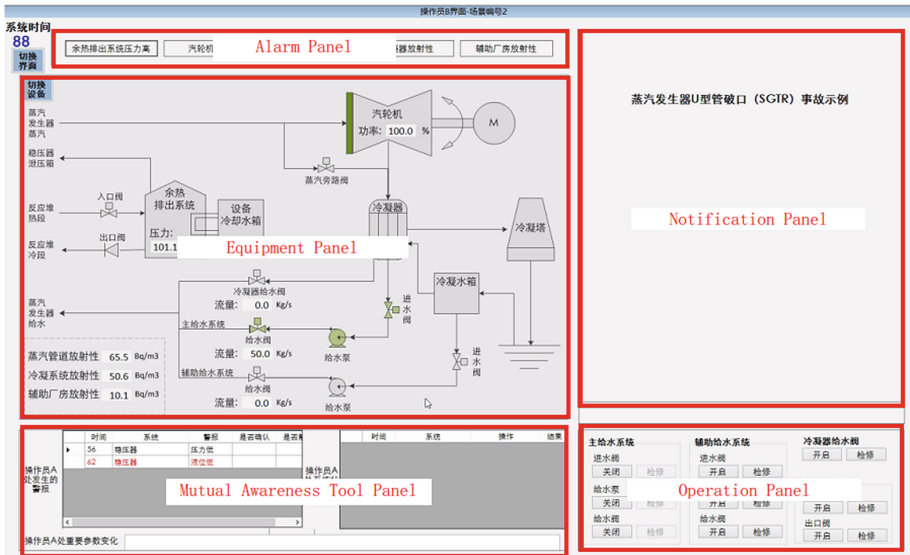


Fig. 2. Interface for conventional island

awareness toolkit, which was first developed by Yuan (2013) [11]. In our study, we did some modification and added a speech display.

The interface of each operator consists of five parts: the *alarm panel* displays an alarm with red alarm tile as any system parameter exceeds its normal range; the

*equipment panel* displays all the equipment and parameters, with an interface-switching button to switch to the other team member's interface; the *notification panel* displays notifications and the emergency operation procedures (EOPs); the *operation panel* provides some buttons that can be clicked to control the equipment; the *mutual awareness tool panel* displays the alarms, operations and important parameter changes of the other team member.

As mentioned above, the participants were divided into two groups, one using system with speech display and the other using system without speech display. The two systems are the same except that the interface with speech display include speech display of the alarms, operations and important parameter changes of the other team member. The voice is Microsoft lili Chinese female voice. It is developed with Server Application Programming Interface (SAPI) and Text to Speech (TTS) provided by Microsoft Speech SDK.

### 3.3 Scenarios and Tasks

There were four type of tasks in the experiment. *Routine operations* required the participants to complete some routine operations according to the notifications. *Alarm monitoring* required the participants to click the alarm tiles when it turned red. *EOP implementation* required the participants to complete the emergency operating procedures (EOPs) when an accident occurred. These three tasks were individual tasks that were finished by oneself without communication. *Accident diagnosis* was a collaborative task and two team members discussed to diagnose accident and decide how to deal with it.

Every team completed three exercise scenarios and two formal scenarios. Every scenario involved several tasks mentioned above and the participants always had to monitor the parameters in all scenarios.

The exercise scenarios were designed to help the participants get familiar with the system. The first exercise was a free exercise, only involving routine operations and alarm monitoring. In the second and third scenarios, an accident occurred and the participants had to perform the EOP and diagnose the accident.

The participants were required to complete all tasks mentioned above and finish some questionnaires in formal scenarios. Two formal scenarios had different complexity. In the low-complexity scenario, loss of coolant accident (LOCA) occurred in the inlet valve of the residual heat removal system. In the high-complexity scenario, loss of heat sink (LOHS) occurred and LOCA occurred in the pressure relief valve of the pressurizer while the participants were performing EOPs of LOHS.

### 3.4 Independent Variables

This was a  $2 \times 2$  mix design with two independent variables.

*Speech display* (with speech display vs. without speech display) is a between-subjects variable. Half of the participants (24 participants, 12 teams) used system with speech display.

*Task complexity* (low-complexity vs. high-complexity) is a within-subjects variable. Every team completed two scenarios with different task complexity. In the low-complexity scenario, only one accident occurred. The team needed to diagnose the accident after the EOPs. In the high-complexity scenario, another accident occurred while they were performing the EOPs of the first accident. The team needed to diagnose the accident and decide whether to take a certain action.

### 3.5 Dependent Variables

There were seven dependent variables.

*Mutual Awareness Score.* In each formal scenario, the participants finished a questionnaire designed with SAGAT method to evaluate the participant's knowledge and understanding of the other team member's status. From the questionnaire we obtained the individual mutual awareness scores. The team's mutual awareness score is the average of the individual scores of the two team members.

*Individual Situation Awareness Score.* We were interested in whether speech display would cause a loss in individual situation awareness. Like mutual awareness score, individual situation awareness score was measured with a questionnaire and we took the average as the final score.

*Alarm missing Rate.* It is the percentage of alarms that were missed by the participants. This was recorded by the system to evaluate individual performance.

*Average Reaction Time.* It is the average time from the appearance of an alarm to the confirmation of the alarm. This was recorded by the system to evaluate individual performance.

*Mental Workload.* In each formal scenario, the participants finished a NASA Task Load Index (NASA-TLX) questionnaire. The participants were required to use 1–10 to measure their mental demand, physical demand, temporal demand, performance, effort, and frustration level. The overall mental workload was the sum of the scores in the six dimensions.

*Diagnosis Score.* It measures the accuracy of the participants' diagnosis. The score has four levels(0, 5, 10, 15), higher if the diagnosis conclusion was closer to the correct answer.

*Analysis Score.* This measure reflects how well the diagnosis was performed by counting how many key issues were addressed in the discussion during the diagnosis process. The percentage of key issues mentioned by the participants was transformed to a 0–10 score.

### 3.6 Procedure

The experiment lasted about 2.5 h. First, the experimenter introduced the experiment to the participants. Then the participants were given a one-hour training on the principle and

operation of the simulated nuclear power plant as well as a quiz to ensure that they mastered what they were taught. After the quiz, they completed the three exercise scenarios in about 40 min. Then they completed the two formal scenarios in about 50 min. The order of the formal scenarios were randomized. During each scenario, the system froze at some time and a questionnaire was given to the participants to measure mutual awareness and individual situation awareness. Then the scenario continued and they perform the EOPs and diagnose the accident. After the diagnosis, mental load was assessed.

## 4 Results

### 4.1 Mutual Awareness

The effects of speech display, task complexity, and their interaction on mutual awareness score were assessed using the repeated measure ANOVA at the significance level of 0.05. There was no significant effect of speech display on mutual awareness score. Mutual awareness score in the low-complexity task was significantly higher than that in the high-complexity task ( $F = 41.050$ ,  $p < 0.001$ ). There was a significant interaction effect between speech display and task complexity ( $F = 4.826$ ,  $p = 0.039$ ).

To further analyze the effect of speech display on mutual awareness score, a t-test was conducted at both task complexity levels respectively. Mutual awareness score of teams with speech display was marginally significantly higher than that of teams without speech display in the low-complexity task ( $t = 1.841$ ,  $p = 0.080$ ). There was no significant effect in the high-complexity task.

### 4.2 Individual Situation Awareness

Homogeneity of variance was rejected using Box test and Levene test at significance level 0.05. The effect of speech display on individual situation awareness score was assessed at both task complexity levels respectively using Mann-Whitney U test at the significance level of 0.05. There was no significant effect of speech display on individual situation awareness score at both task complexity levels. The effect of task complexity on individual situation awareness score was assessed when team used a system with and without speech display respectively using Wilcoxon signed-rank test. Individual situation awareness score in the low-complexity task was significantly higher than that in the high-complexity task whether a team used a system with ( $p = 0.03$ ) or without ( $p = 0.03$ ) speech display.

### 4.3 Alarm Missing Rate

There was no alarms at the conventional island in the high-complexity scenario and thus we only analyzed alarm missing rate in the low-complexity task. Normality was rejected using Anderson-Darling test at the significance level of 0.05. The effect of

speech display on alarm missing rate in the low-complexity task was assessed using Mann-Whitney U test at the significance level of 0.05. No significant effect was found.

#### 4.4 Average Reaction Time

Similarly, average reaction time was analyzed only for the low-complexity task. Normality was rejected using Anderson-Darling test at the significance level of 0.05. The effect of speech display on average reaction time in the low-complexity task was assessed using Mann-Whitney U test and no significant effect of speech display was found on average reaction time.

#### 4.5 Mental Workload

The effects of speech display, task complexity, and their interaction on the overall mental workload were assessed using the repeated measure ANOVA at the significance level of 0.05. There were no significant main effects and interaction effects.

However, when we analyzed each dimension of mental workload separately, we found that the performance dimension of the teams with speech display was significantly higher than the teams without speech display ( $F = 5.576$ ,  $p = 0.022$ ). The performance dimension of mental workload measures the extent that the participants thought they had performed the tasks in the experiment well. The teams with speech display inclined to think that they had not performed the tasks well.

#### 4.6 Diagnosis Score

Diagnosis score was a four-level ordered categorical variable. The effects of speech display and task complexity on diagnosis score were assessed using a Logistic model. Diagnosis score in the low-complexity task was found to be significantly higher than that in the high-complexity task ( $Chi-square = 24.799$ ,  $p < 0.001$ ). There was no significance effect of speech display on diagnosis score.

#### 4.7 Analysis Score

The effects of speech display, task complexity, and their interaction on analysis score were assessed using the repeated measure ANOVA at the significance level of 0.05. There were no significant effect of speech display on analysis score. Analysis score in the low-complexity task was significantly higher than that in the high-complexity task ( $F = 19.181$ ,  $p < 0.001$ ). There was a marginally significant interaction effect between speech display and task complexity ( $F = 2.952$ ,  $p = 0.100$ ).

To further assess the interaction effect, a t-test was conducted on the effect of speech display on analysis score at both task complexity levels, respectively. We found that analysis score of the teams with speech display was significantly lower than that of



the teams without speech display in the low-complexity task ( $t = -2.894$ ,  $p = 0.008$ ). There was no significant effect in the high-complexity task.

## 5 Discussion and Further Plan

According to the results, speech display improved mutual awareness only in the low-complexity task. One possible reason is that information load in the high-complexity task was so high that no more attention could be paid to the other team member even with speech display. Moreover, we found that the performance dimension of mental workload of teams with speech display was higher. One possible reason is that the participants realized that they missed some important information as they could not remember all the information displayed by speech. Thus, they thought they had not performed well in the experiment. We also found that analysis score of the teams with speech display was lower in the low-complexity task. This indicates that speech display had an adverse effect on performance to some degree.

In this study, we analyzed the effect of speech display in mutual awareness tools on team mutual awareness, individual performance, mental workload and team performance on two levels of task complexity. For the next stage, tone without speech could be used and the effect of this kind of auditory display could be examined.

**Acknowledgements.** This work was supported by the National Natural Science Foundation of China (Grant No. 71371104) and the Open Funding Project of National Key Laboratory of Human Factors Engineering (Grant No. HF2012-K-02).

## References

1. Gutwin, C., Greenberg, S.: Effects of awareness support on groupware usability. In: Proceedings of International Conference on Human Factors in Computing System, pp. 511–518 (1998)
2. Schmidt, K.: Some notes on mutual awareness, COTCOS-Report, Universidade Técnica da Dinamarca, disponível em <http://citeseerist.psu.edu/330273.html>, último acesso em Jan 1998
3. Oviatt, S., Coulston, R., Lunsford, R.: When do we interact multimodally? cognitive load and multimodal communication pattern. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 129–136 (2004)
4. Sarter, N.: B.: Multimodal information presentation in support of human-automation communication and coordination. *Adv. Hum. Perform. Cogn. Eng. Res.* **2**, 13–35 (2002)
5. Dowell, J., Shmueli, Y., Kingdom, U.: Blending speech output and visual text in the multimodal. *Interface* **50**(5), 782–788 (2008)
6. Kalyuga, U., Chandler, P., Sweller, J.: cc in Multimedia Instruction, vol. 371, April 1998, pp. 351–372 (1999)
7. Moreno, R., Mayer, R.E.: Learning science in virtual reality multimedia environments: role of methods and media. *J. Educ. Psychol.* **94**(3), 598 (2002)
8. Wickens, C.D.: Multiple resources and performance prediction. *Ergon. Psychol. Mech. Model. Ergon.* **83** (2005)

9. Wickens, C.D., Liu, Y.: Codes and modalities in multiple resources: a success and a qualification. *Hum. Factors J. Hum. Factors Ergon. Soc.* **30**(5), 599–616 (1988)
10. Horrey, W.J., Wickens, C.D.: Driving and side task performance: the effects of display. *Hum. Factors.* **46**(4), 611–624 (2004)
11. Xihui, Y., Zhizhong, L.: Effects of mutual awareness on team performance in accident diagnosis under computerized working environment. Dept. of Industrial engineering, Tsinghua University (2015)