# Experimental Study of Indices for Evaluating the Augmented Reality System from the Viewpoint of Human Centered Design

Hidenori Fujino[1], Hirotake Ishii[1] and Hiroshi Shimoda[1]

[1]Graduate School of Energy Science, Kyoto University, Kyoto, Japan
(Tel: +81-774-38-4403, E-mail: fujino@uji.energy.-u.ac.jp)

**Abstract**: Applying augmented reality to a work support system is expected to decrease human errors when practicing the maintenance of equipments of nuclear power plants. It is necessary to establish the indices to evaluate the system from the viewpoint of human centered design in order to develop such a system. In this study, therefore, the authors investigated the applicability of some indices to evaluate the system. As a result, it was revealed that task completion time and the score of NASA-TLX are appropriate indices.

**Keywords:** Augmented Reality, Indices for evaluating, Display time delay, Human centered design

## 1. INTRODUCTION

The area of nuclear power plant is very large and the structure is very complex. Moreover there are very large amount of machines of facilities in there and each machine consists of very large amount of parts. Therefore, when practicing the maintenance work of nuclear power plant, using the work support system applied augmented reality is expected to decrease human error, which gives the information about work to workers intuitively; which machine is to be maintained, where the machine to be maintained is, or how the machine is to be maintained. Example image of this idea is shown in Fig.1. In this figure, one bulb is pointed out in the scene that there are some same shape bulbs. That bulb is the target of operation. By augmented reality, such an intuitive information presentation can be realized and the error of operating wrong bulb would be decreased.

It is necessary to develop various elementary systems in order to develop such a work support system; tracking system, calibration system, registration system, situation recognition system and so on [1] and also it is necessary to fix requirements to system performance to develop these systems actually. Requirements to system performance should be fixed based on usability test with mock-up models or prototypes in a real scene [2]. At present, however, such a usability test is not held practically. Moreover, indices to evaluate the system from the view point of human centered design have not been established even yet.

In this study, therefore, the author aimed at establishing such an evaluation indices of the augmented reality system.

## 2. SURVEY AND DISCUSSION OF INDICES FOR THE EVALUATION

### 2.1 Survey of indices

Generally speaking, the influences to a user using a work support system with augmented reality is considered to consist of three aspects; the influence to user's task performance, the influence to user's psychological aspect, and the influence to user's physiological aspect. Then, indices to evaluate the system from these three points used on not only the field of augmented reality but also the other various fields are listed up as follow:

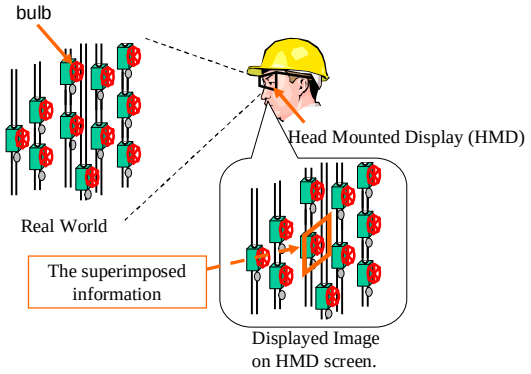


Fig. 1 The example image of work support system applied augmented reality

"Task completion time" and "possibility of task errors" are generally used as indices to evaluate the influence to user's task performance.

"Mental work load by NASA-TLX [3]" and "Subjective Feeling of Fatigue [4]" are used to to evaluate the influence to user's psychological aspect generally.

"Electromyogram [5]", "brain waves [5]", "heart beat rate [5]", "critical flicker frequency [6] "urinary adrenaline [7] "salivary chromogranin [8]" and "catecholamine from blood sample [8]" are used to evaluate the influence to user's physiological aspect.

### 2.2 Discussion of indices to evaluate the work support system with augmented reality

There are 3 requirement points for the indices to evaluate the work support system with augmented reality:

1. There should be only a few devices to take on a user's body for measurement, and those devices

should not disturb user's action for his work.

2. Physical or mental stress given to a user in measuring should be small.
3. Time for measurement should be short in order to prevent that the influence be lost.
4. It is unnecessary for a user to be skilled for measuring the indices.

The applicability of each index listed up above was considered based on these requirements. As a result of the consideration, electromyogram, brain waves, heart beat rate, catecholamine from blood sample, urinary adrenaline, and critical flicker frequency are considered not to be applicable. The reasons are as follow;

- Electromyogram, brain waves, and heart beat rate require a user to take many probes and lines on their body. Those probes and lines would disturb the user's action for his work. Moreover, to take such devices on itself would give any stress to the user.
- Catecholamine from blood sample requires sampling a user's blood with a syringe. That would give mental stress to the user.
- Urinary adrenaline requires a little time to take a sample from a user. Therefore, the influence of using the system might be lost at taking a sample.
- Critical flicker frequency requires some skills to a user for measuring the value precisely.

According to the consideration, task completion time, possibility of task errors, mental work load by NASA-TLX, subjective feeling of fatigue, and salivary chromogranin, are considered the applicable indices to evaluate the work support system with augmented reality.

In the next section, the subject experiment to confirm the actual applicability of each index of these is explained.

## 3. SUBJECT EXPERIMENT

### 3.1 The purpose of the experiment.

Since the list of previous section of indices is based on only a consideration from the viewpoint of the fulfillment of requirements to evaluate the work support system, it is necessary to investigate the actual sensitivity of each index against the change of performance of the work support system. Therefore, the purpose of the experiment is to do the investigation and to discuss the applicability of each index as the index to evaluate the work support system with augmented reality.

### 3.2 The method of the experiment

The method for evaluating the sensitivity of each index in the experiment is as follow:

Some experimental trials which use the augmented reality system with different performance in "display time delay" each other are practiced. Then, one experimental trial is to measure each index after user's doing the experimental task with receiving the work support information by the augmented reality system. After that, the applicability of each index is discussed based on the value of each index measured in each trial from the viewpoint of the effect of performance difference of the augmented reality system used in each trial.

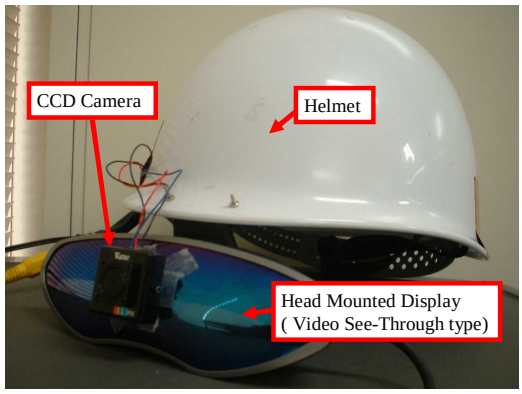


Fig. 2 The hardware of developed system

### 4.3 The system for the experiment

The work support system with augmented reality was developed for the experiment applied with ARToolKit [9]. The hardware of developed system is shown in Fig.2. In this system, the length of display time delay can be set freely. Display time delay is the delay from capturing the image by CCD camera to displaying the superimposed information on the head mounted display. Note that only the superimposed information is delayed and a video picture of environment captured CCD camera is displayed on the screen of HMD on real time like optical see-through type since virtual reality sick would be happened and affect indices together with display time delay if a video picture also be delayed. Because the purpose of this experiment is to investigate the sensitivity of each index against the just change of the performance of one point, that is display time delay, other factors which affect indices should be eliminated. The computer for information processing is not on a subject but is held by the conductor of this experiment. Therefore, the conductor stands by a subject and moves with him.

### 3.4 The task for subjects in the experiment

The task for subjects in the experiment is "a target search", which is to find and touch the 30 target pictures on the wall one by one according to the direction of the system with augmented reality explained in 4.3. In that system, a sphere is described on the target picture if the target picture to be touched is in the captured picture, and a quadrangular pyramid is described on one marker near the center of the captured picture if the target is not in. The tip of the quadrangular pyramid directs the place of the target to a subject.

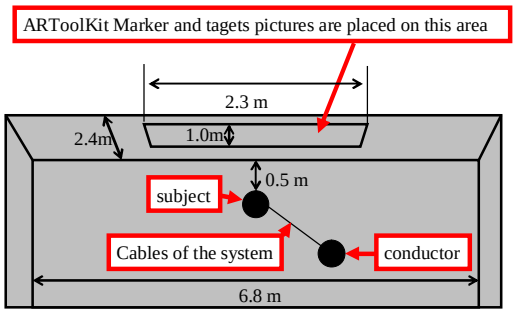


Fig. 3 The bird's-eye view of the experimental environment.

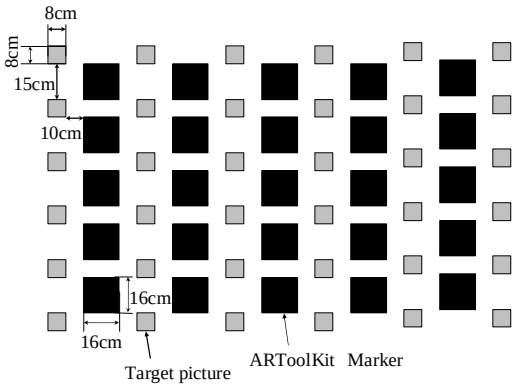


Fig. 4 The placement of ARToolKit markers and target pictures

The bird's-eye view of the experimental environment is shown in Fi.3 and the placement of ARToolKit markers and target pictures is shown in Fig.4. Shown in Fig.3, a subject was told to stand 0.5m away from the wall and was allowed only to move crossly.

### 4.5 Subjects

The number of subjects was 6 and the oldest subject was 23 years old and the youngest was 21. All subject have never experienced to use any augmented reality system.

### 4.6 Experimental condition, order, work flow

5 conditions of the performance of display time delay were prepared, which are 33ms, 100ms, 166ms, 300ms, and 700ms. And 11 patterns of the order of the targets to be directed by the system were prepared. 1 pattern of 11 was for the trial to experience the experimental system and other 10 patterns were for practice. Each pattern of the order was different from others, but the length of the line of flow in each pattern was same each other, which was measured according to the assumption that a subject moves along the ideal path. Each subject engaged in all 10 patterns of the target search according to the directions by the system which was changed the performance of display time delay as table 1.

The work flow of each subject in the experiment

Table 1 The experimental order of each subject

(unit: ms, S: Subject, Pat: Patter of the order of direction)

| | S.A | S.B | S.C | S.D | S.E | S.F |
|---|---|---|---|---|---|---|
| Pat.1 | 33 | 33 | 33 | 33 | 33 | 33 |
| Pat.2 | 100 | 100 | 100 | 100 | 100 | 100 |
| Pat.3 | 166 | 166 | 166 | 166 | 166 | 166 |
| Pat.4 | 300 | 300 | 300 | 300 | 300 | 300 |
| Pat.5 | 700 | 700 | 700 | 700 | 700 | 700 |
| Pat.6 | 33 | 33 | 33 | 33 | 33 | 33 |
| Pat.7 | 166 | 300 | 100 | 100 | 166 | 300 |
| Pat.8 | 100 | 100 | 300 | 166 | 300 | 166 |
| Pat.9 | 300 | 166 | 166 | 300 | 100 | 100 |
| Pat.10 | 700 | 700 | 700 | 700 | 700 | 700 |

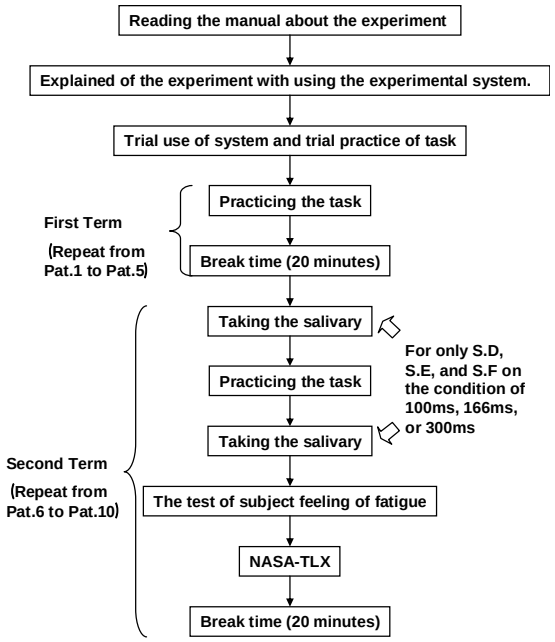


Fig.5 The work flow of each subject in the experiment

was shown in Fig.6. The test of salivary chromogranin was practiced for only 3 subjects; subject D, subject E, and subject F.

Note that all subject experienced all 5 conditions of the performance of display time delay in trial use of system for the purpose of avoiding that the effect of embarrassment to firstly experience the system under each condition would affect to the result.

### 4.7 Result

#### Task completion time

The result of task completion time of each subject on each pattern and each term is shown in Fig.6 – Fig.11.

On all subjects, the differences of task completion time of each display time delay between in first term and in second were almost same. Therefore, it is considered that the effect of learning was small enough to be negligible.

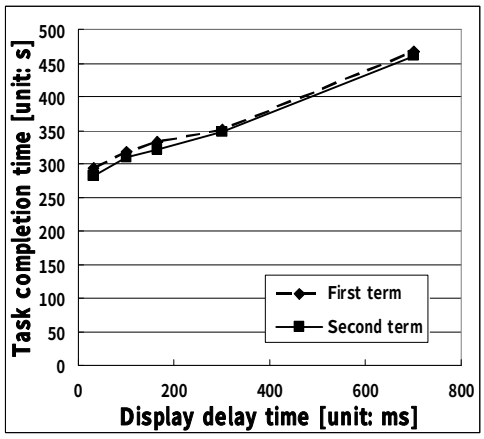


Fig. 6 The result of task completion time of Subject A on each display delay time and each term

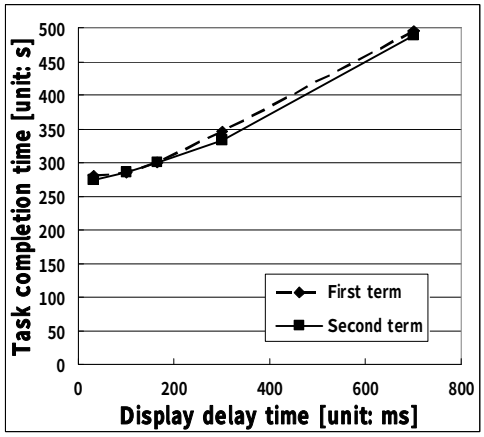


Fig. 7 The result of task completion time of Subject B on each display delay time and each term

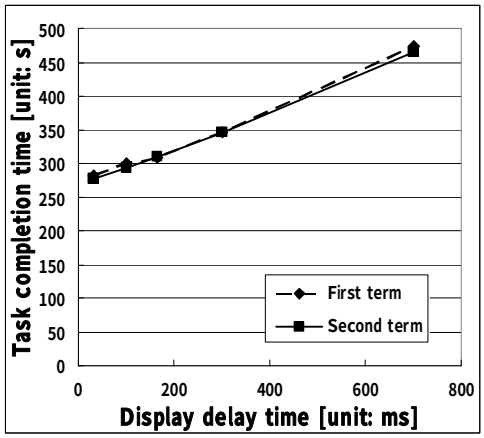


Fig. 8 The result of task completion time of Subject C on each display delay time and each term

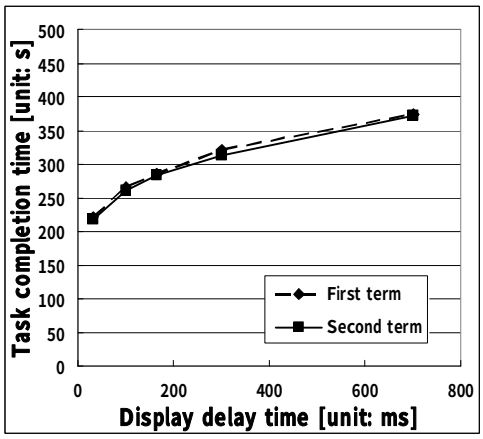


Fig. 9 The result of task completion time of Subject D on each display delay time and each term

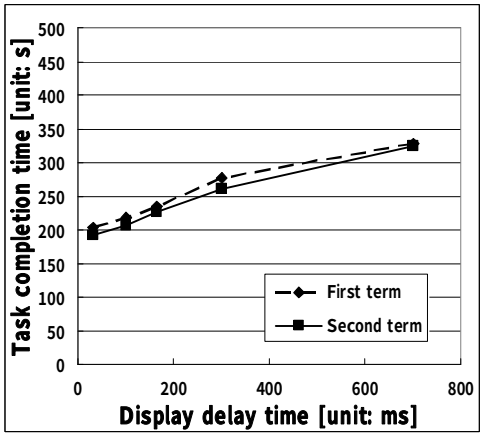


Fig. 10 The result of task completion time of Subject E on each display delay time and each term

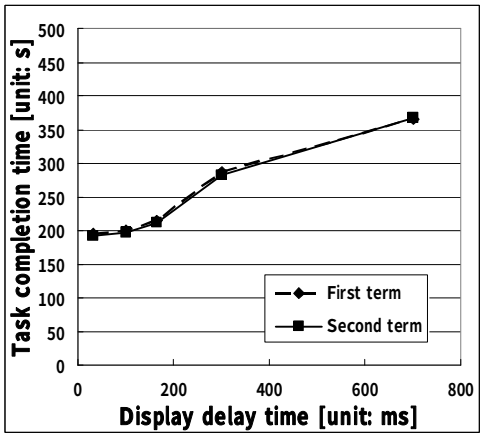


Fig. 11 The result of task completion time of Subject F on each display delay time and each term

As the results shown in Fig.6 – Fig.11, on all subjects, task completion time became longer as display time delay was set longer. Especially, task completion time became longer even if there was little difference such as between 33ms and 100ms. Moreover, on all subjects, the expansions of task completion time between 33ms and 100ms were larger than the expansion for the difference of the processing speed between each system (the expansion for the difference of the processing speed between each system was considered of 4.02 [sec] for calculation.)

As th e result of these, it can be considered that task completion time is sensitive enough to evaluate the system.

**Possibility of task errors**

Subject A made one error on 300 ms and each of subject A, subject B, and subject C made one error on

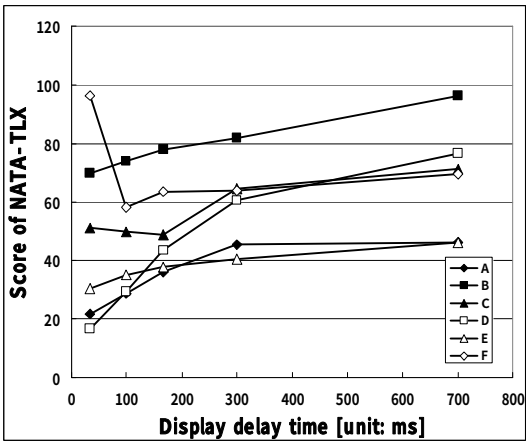

Fig. 12 The result of the score of NASTA-TLX of each subject on each display delay time

700 ms. Therefore, since the number of all error was only 4, it can be considered that the sensitivity of possibility of task errors is not enough. However, since the number of error became more as display time delay was longer, the sensitivity may be increased if more complex and more difficult task would be prepared.

**NASA-TLX**

The result of NASA-TLX is shown in Fig.12. The score of NATA-TLX of each subject had a tendency to become higher as display delay time became longer. However, on subject C and subject F, the score was down in the range that display time delay was short. Therefore, it is considered that NASA-TLX can be used to evaluate the influence to a user, but it is necessary to conduct more precise experiment about the sensitivity of NASA-TLX.

**Subject feeling of fatigue**

The result of the test of subject feeling of fatigue of each subject is shown in Fig.13 – Fig.18. As these result, it can be seen that there was no consistency on subject feeling of fatigue. The reason of the result is considered that subject's fatigue be very small since prepared task might be not so complex and difficult. Therefore, it is necessary to investigate if the consistency would appear when more complex and difficult task would be prepared.

**Salivary chromogranin**

The result of measurement of salivary chromogranin is shown in Table.2. As the result, on each display time delay and on each subject, the difference between before task and after task was very small and there was not a consistency in a tendency of changing the value. Therefore, it is considered that salivary chromogranin was not useful to evaluate the influence to a user.

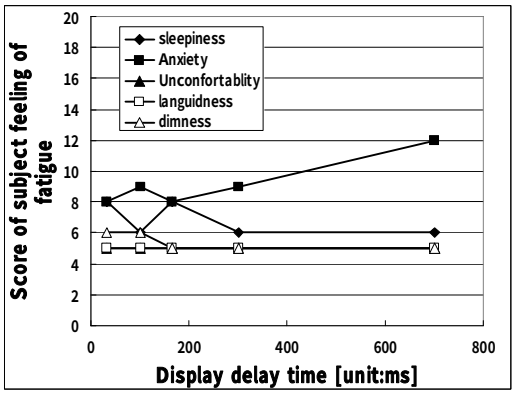

Fig. 13 The result of subject A's feeling of fatigue on each display delay time

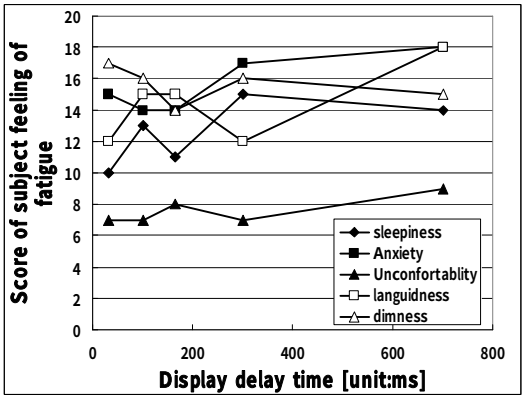

Fig. 14 The result of subject B's feeling of fatigue on each display delay time

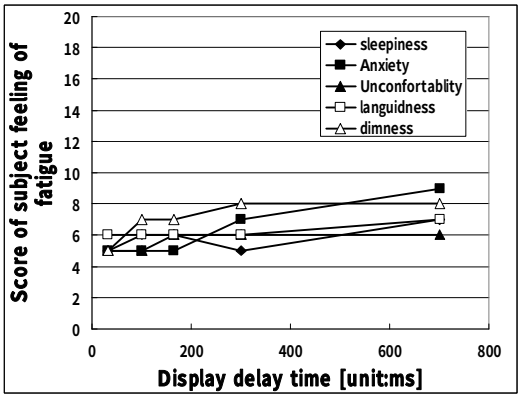

Fig. 15 The result of subject C's feeling of fatigue on each display delay time

## 5. CONCLUSION

In this study, the author aimed at establishing evaluation indices of the augmented reality system from the viewpoint of human centered design. According to literary survey and consideration of the feature of augmented reality system, task completion time, possibility of task error, NASA-TLX, subjective fatigue and salivary chromogranin were picked up as the candidate indices to evaluate augmented reality system.

For confirming the sensitivity of each index, a subject experiment was conducted. As the result of

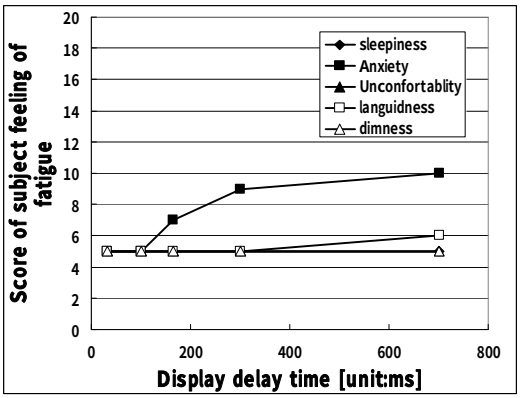

Fig. 16 The result of subject D's feeling of fatigue on each display delay time

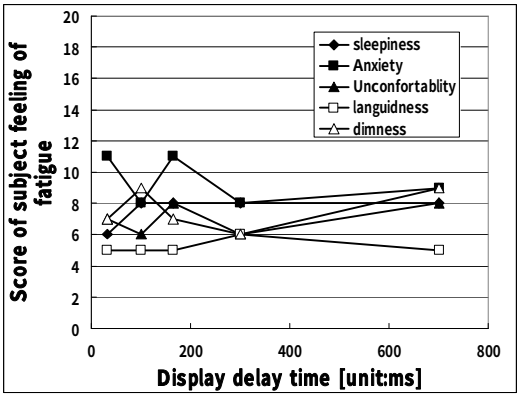

Fig. 17 The result of subject E's feeling of fatigue on each display delay time

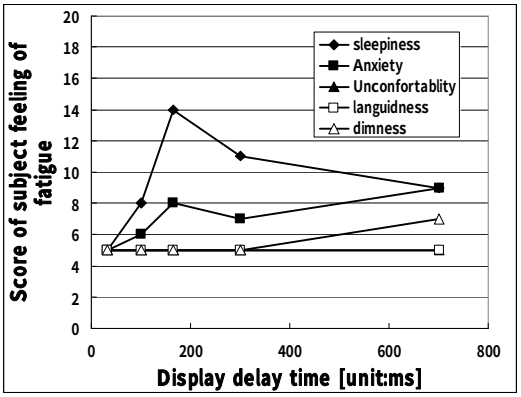

Fig. 18 The result of subject D's feeling of fatigue on each display delay time

the experiment, the task completion time and the score of NASA-TLX increased along with the increase of the display time delay, and there were no significant variation of the other indices. This means that task completion time and NASA-TLX are appropriate indices to evaluate augmented reality system.

Since the experiment conducted in this study is limited for only one fixed task and only one independent value of experimental condition, it is necessary to expand the task field and performance factors to establish the evaluation indices of augmented reality systems.

Table 2 The result of measurement of salivary chromogranin

| display time | Taking salivary | Subject D | Sbuject E | Subject F |
|---|---|---|---|---|
| 100 ms | before practicing | 0.96 | 1.24 | 1.90 |
| | after practicing | 1.14 | 1.15 | 1.74 |
| 166ms | before practicing | 1.40 | 1.35 | 1.85 |
| | after practicing | 1.40 | 1.54 | 2.00 |
| 300ms | before practicing | 1.26 | 1.34 | 1.69 |
| | after practicing | 1.25 | 1.18 | 2.06 |

## REFERENCES


[1] R. T. Azuma, "A Survey of Augmented Reality", Presence: Teleoperators and Virtual Environments, Vol.6, No.4, pp.355-385, 1997.

[2] ISO, ISO-13407:1999 "Human-centred design processes for interactive systems".

[3] A. Tang, C. Owen, F. Biocca, W. Mou, "Experimental Evaluation of Augmented Reality in Object Assembly Task", Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02), pp.265-266, 2002.

[4] http://square.umin.ac.jp/of/ (12th., May, 2007) (in Japanese)

[5] H.Uchiyama, Master Thesis: "Physiological evaluation of PC input interface", Graduate School of Electronics, Tokyo Denki University, 2004. (in Japanese)

[6] H.Ichikawa, K.Honda, M.Ohashi, "Relationship between VDT Operation, and Stress and Fatigue - In case of Telework and E-learning –",The Journal of The Infosocionomics Society, Vol.1, No1, pp.64-72, 2006.(in Japanese)

[7] T.Taguchi, "Evaluation of Fatigue during Car Driving", The R&D Review of Toyota CRDL, Vol.33, No.4, pp.25-31, 1998. (in Japanese)

[8] H.Nakane, "Salivary Chromogranin A as Index of Psychosomatic Stress Response", The R&D Review of Toyota CRDL, Vol.34, No.3, pp.17-22, 1999. (in Japanese)

[9] http://www.hitl.washington.edu/artoolkit/ (12th., May, 2007)