

# Proper Presentation Order of the Contents in Voice Guidance When Operating Machine According to Guidance

Masato Sakurai\*, Shintaro Tahara\*\*, and Sakae Yamamoto\*\*

\**Kanazawa Institute of Technology, masato@neptune.kanazawa-it.ac.jp*

\*\**Tokyo University of Science*

**Abstract:** To examine the effect of the presentation order of contents when pressing the buttons according to the voice guidance the suitability of voice guidance was evaluated varying presentation order of contents in the voice guidance. The presentation order of three contents (place, what, and action contents) included in voice guidance was replaced. The task was to press the button on the touch panel according to the instruction of voice guidance. In the experiment, the subjective evaluations, reaction time, and path length of eye movement during the task were measured. As the results, when the voice guidance is composed that “what content” is given to the next of “place content” in the presentation order it is evaluated highly in the measurements. It means that the presentation order of contents in voice guidance influences the subjective evaluations and performance of the task when pressing the button according to the voice guidance.

**Key words:** *voice guidance, presentation order, contents, subjective evaluation, eye movement*

## 1. Introduction

Recently, many products and equipments are installed the voice guidance to support when users operate it. The voice guidance is the instruction with voice of supporting to make users obtain the results they want [1]. It makes users pay attention to its operation when using the product or equipment. It does not need to have any visual space (i.e. displaying space) to present the voice guidance, and it is not necessary to turn their eyes to the voice guidance. Therefore, it will be applicable in many areas in future.

There is little study for investigating the effects of the contents in voice guidance although the contents are one of the important factors for the voice guidance as well as the physical characteristics. From the survey of the contents in the product installed the voice guidance, it consists of three contents. One of them indicates the place where to operate when users use it (i.e. place content). The other one represents what to operate (i.e. what content). The last one is how to operate as action (i.e. action content). There is no study for the proper presentation order of the above three contents in voice guidance when user uses the machine. In this study, the purpose is to examine the effect of the presentation order of contents when pressing the buttons according to the voice guidance varying presentation order of contents. In the experiment, the subjective evaluation was measured in terms of ease of understanding and the nature of Japanese language for the voice guidance. The path length of eye movement during the task and the reaction time were also measured.

## 2. Experiment

### 2.1 Task

The task was to press the button on the touch panel according to the instruction of voice guidance. Fig. 1 shows the screen of the touch panel in this experiment. Prior to the task, the fixation mark was presented on the center of the screen for 2 s. And the subjects were asked to press the proper button on the touch panel with the instruction of voice guidance after the beep. Then, the subjects evaluated subjectively on

7-rank scale (1: strongly disagree, 2: disagree, 3: slightly disagree, 4: neither agree nor disagree, 5: slightly agree, 6: agree, 7: strongly agree) in the following questionnaires; (a) how much did you feel it was suitable for voice guidance? (b) how much did you feel it was easy to understand the voice guidance? (c) how much did you feel it was natural as Japanese language? The presentation order of contents (place, what, and action contents) in voice guidance was randomly changed for each trial. Also, the descriptions of the positions and buttons (A, B, C and “i”, “ro”, “ha” in Katakana of Japanese in Fig. 1) were randomly changed for each trail.

## 2.2 Voice guidance

The presentation order of three contents (place, what, and action contents) included in voice guidance was replaced, and voice guidance of six conditions was created as shown in Table 1. As the explanation of the place content, the voice guidance which directs ‘A’, ‘B’, and ‘C’ was created. As the explanation of the what content, the voice guidance which directs “i”, “ro”, and “ha” in Katakana of Japanese was created for corresponding to each button. As the explanation of the action content, the voice guidance which directs to push a button was only created.

The speech rate of the voice guidance was used six morae/s according to the report of previous study [2]. The volume of sound (sound pressure) in the voice guidance was adjusted by each subject so that it was enough to listen. The voice guidance was made of speech synthesis and its speaker corresponded to a female. Each of the voice guidance was presented one time in each trial.

## 2.3 Procedure

The subject performed the calibration so that a touch panel could be correctly operated before the task. After the calibration, the subject performed the task with the presented voice guidance. Each subject randomly

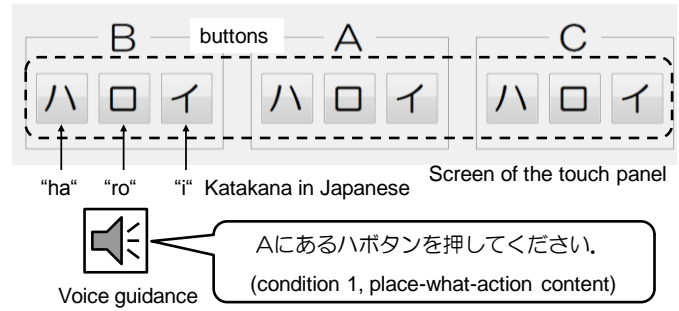


Fig. 1 Screen of the touch panel in the experiment including the voice guidance.

Table 1 Presentation order of contents in the voice guidance in the experiment (conditions).

Condition	order	place – what – action
	Japanese	Aにあるハボタンを押してください。
Condition 2	order	place – action – what
	Japanese	Aにあるボタンを押してください。ハボタンです。
Condition 3	order	what – place – action
	Japanese	ハボタンはAにあります。押してください。
Condition 4	order	what – action – place
	Japanese	ハボタンを押してください。Aにあります。
Condition 5	order	action – place – what
	Japanese	押すのはAにあるハボタンです。
Condition 6	order	action – what – place
	Japanese	押すのはハボタンです。Aにあります。

performed the task for 18 times (for six conditions in the voice guidance, and repeated three times for each condition).

## 2.4 Apparatus

The experiment was run on a tablet PC (lenovo, ThinkPad X200 Tablet). The program was developed in Visual Basic with Visual Studio 2010. The speaker (SONY, SRS-TD60) was used to present the voice, connecting to the tablet PC. The observers sat to the chair and performed the task with their dominant hand using the tablet on the table. The viewing distance was set 500 mm using the chin rest to fix their head, and the visual angles of horizontal and vertical axes in the screen were 14.6 and 9.1 deg, respectively. The camcorder (SONY, Handycam HDR-S12) was used to record the subjects' operation during the task. The reaction time in performing the task was measured from the images recorded by the camcorder for each subject. The reaction time was defined as the time from presenting the voice guidance to pressing the proper button instructed by the voice guidance. Eye mark recorder (nac, EMR-9) was also used to measure the subjects' eye movement during the task.

## 2.5 Subjects

Ten subjects participated in this experiment. The average age was 22.2 years (S.D. 1.3). All the subjects had normal visual acuity and normal hearing.

## 3. Results and Discussions

### 3.1 Subjective evaluation

Fig. 2 shows the results of the subjective evaluation regarding the suitability, ease of understanding, and naturalness of Japanese language for the voice guidance (questionnaires (a)-(c) of 2.1) over all the subject responses for six conditions in the voice guidance. The horizontal and vertical axes indicate each condition in the voice guidance and subjective evaluation value, respectively. Each bar represents the average results of each evaluation including the standard deviation within the subjects.

In this figure, the values of three evaluations in conditions 1 and 5 are relatively higher than the others, especially the condition 1 is the highest in all. From the results of one-way ANOVA for three subjective evaluations, there is a significantly difference of the presentation order of the contents in the voice guidance in this experiment ( $\alpha=0.01$ ). It means that the presentation order of the

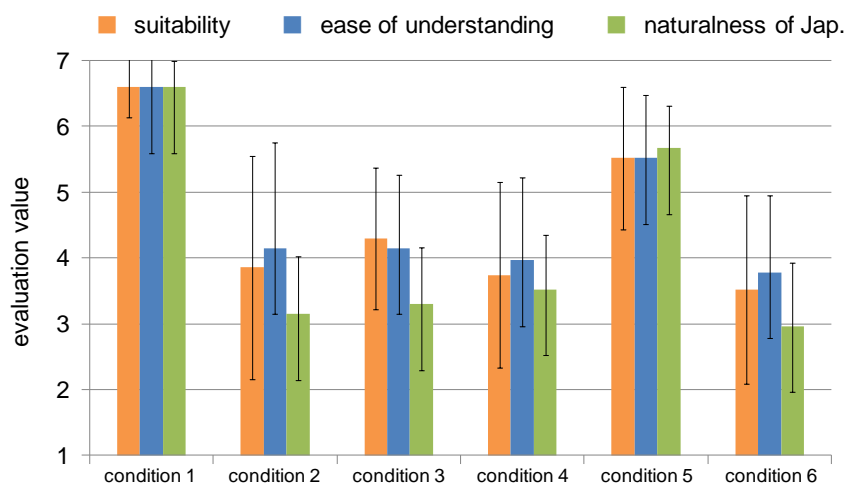


Fig. 2 Average results of the subjective evaluation regarding the suitability, ease of understanding, and naturalness of Japanese language for the voice guidance (questionnaires (a)-(c) of 2.1) over all the subject responses for six conditions in the voice guidance.

contents in the voice guidance influences their evaluations. It obtains the correlation between all the evaluations and it is considered that the ease of understand and naturalness of Japanese language contributes to the suitability of the voice guidance in subjective evaluation.

In terms of a feature for the explanation of the voice guidance, the conditions 1 and 5 are composed that “what content” is given to the next of “place content”. It can search the proper button removing the information from the presented voice guidance. It means that it can reduce as much as possible the load in working memory [3]. And, the voice guidance of condition 1 and 5 can be better presentation order of contents in terms of ease of understanding in voice guidance. Also, the conditions 1 and 5 consist of one sentence in Japanese and it is not necessary to divide into some sentences such as other conditions. It means that they are natural in Japanese language. Therefore, the presentation order of contents in conditions 1 and 5 is suitable when operating machines according to the voice guidance.

### 3.2 Reaction time

Fig. 3 shows the results of the reaction time in the experiment over all the subjects' responses for six conditions in the voice guidance. The horizontal and vertical axes indicate each condition in the voice guidance and the reaction time, respectively. Each bar represents the average results including the standard deviation within the subjects.

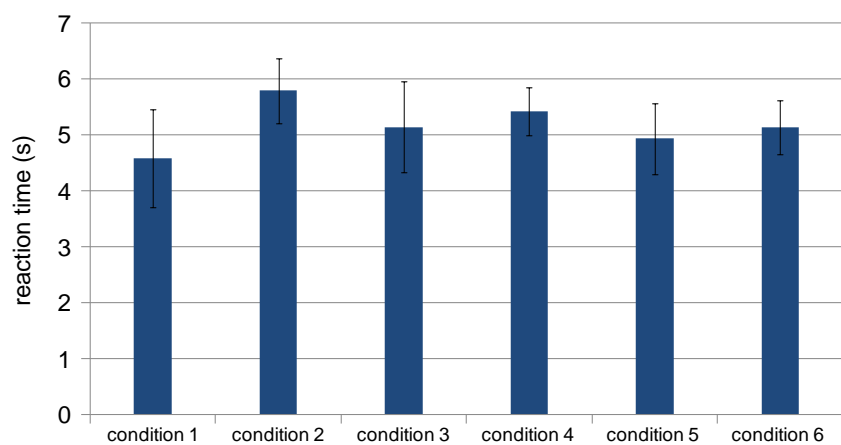


Fig. 3 Average results of the reaction time in the experiment over all the subjects' responses for six conditions in the voice guidance.

The reaction time of conditions 1 and 5 are relatively shorter than those of other conditions. From the results of one-way ANOVA for the reaction time, there is a significantly difference of the presentation order of the contents in the voice guidance in this experiment ( $\alpha=0.01$ ). Compared with the results of subjective evaluation, there is a tendency that the conditions highly evaluated become a short reaction time. It is suggested that the reaction time contributes to the evaluation of suitability in the voice guidance. Condition 1 is the shortest reaction time and it consists of the presentation order of “place-what-action content” in the voice guidance. Its presentation order can make subjects operate it quickly as much as possible compared with those in other conditions. Its reaction time is the shortest as the results and it is an efficient way to operate machines according to the voice guidance.

### 3.3 Path length of eye movement

Fig. 4 shows how to calculate the path length of eye movement in horizontal direction during the task. From the coordinate data measured by the eye mark recorder for each subject, it sums the path length of eye movement in horizontal direction during task. It corresponds to the total of the lines in the bottom of this figure.

Fig. 5 shows the results of the ratio of path length of eye movement in horizontal direction over all the subjects' measurement for six conditions in the voice guidance. The path length represents the ratio for that of condition 1

in each subject. The horizontal and vertical axes indicate each condition in the voice guidance and the ratio of the path length, respectively. Each bar represents the average results including the standard deviation within the subjects.

The ratio of path length in conditions 1 and 5 are smaller than those of other conditions. From the results of one-way ANOVA for the ratio of path length, there is a significantly difference of the presentation order of the contents in the voice guidance in this experiment ( $\alpha=0.01$ ). Compared with the results of subjective evaluation, there is a tendency that the conditions highly evaluated become relatively a small path length of eye movement in horizontal direction. It is suggested that the path length of eye movement contributes to the evaluation of suitability in the voice guidance. In

terms of a feature for the presentation order of contents in the voice guidance, the conditions 1 and 5 are composed that “what content” is given to the next of “place content” as shown in the above. It means that the area of visual searching for the button selection is narrow when obtaining “place content” ahead compared with obtaining “what content” ahead for the voice guidance instructed. Therefore, its path length of eye movement in horizontal direction is shorter than that of others. And it is related to reducing as much as possible for the load in working memory when searching the button according to the voice guidance as the above mentioned. Overall, the presentation order of contents in conditions 1 and 5 is suitable when operating machines according to the voice guidance.

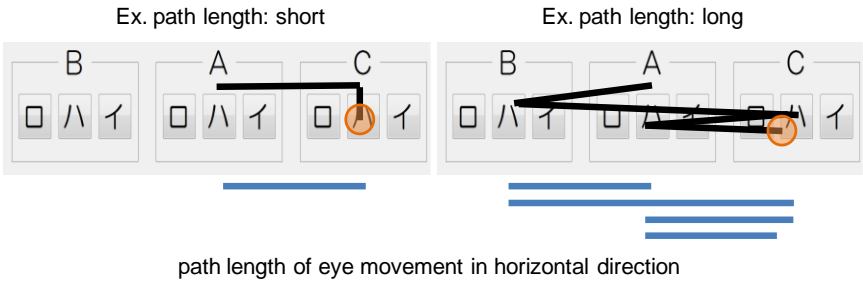


Fig. 4 How to calculate the path length of eye movement in horizontal direction during the task.

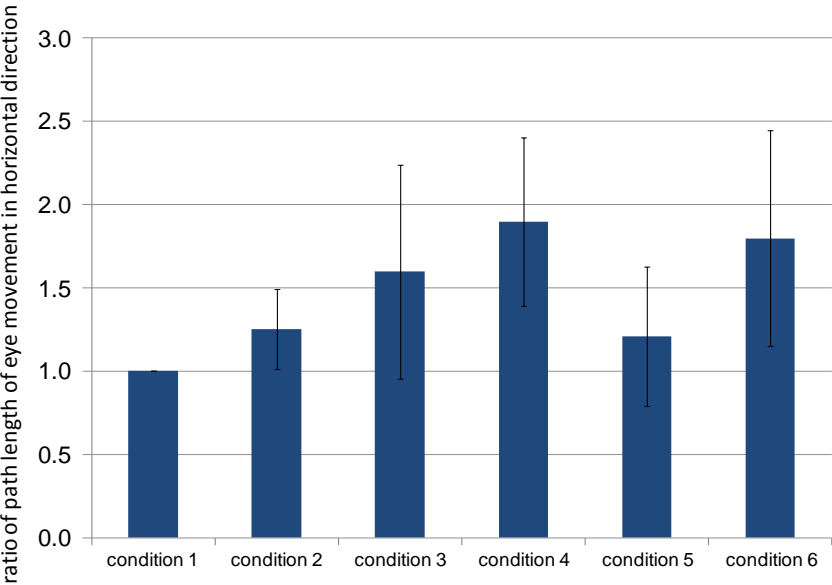


Fig. 5 Average results of the ratio of path length of eye movement in horizontal direction over all the subjects' measurement for six conditions in the voice guidance. The path length represents the ratio for that of condition 1 in each subject.

#### 4. Conclusion

The suitability of voice guidance when pressing the button according to the voice guidance was evaluated varying the presentation order of contents in the voice guidance. The presentation order of three contents (place, what, and action contents) included in voice guidance was replaced, and voice guidance of six conditions was created. The task was to press the button on the touch panel according to the instruction of voice guidance. In the experiment, the subjective evaluation was measured in terms of suitability, ease of understanding and the naturalness of Japanese language for the voice guidance. The path length of eye movement during the task and the reaction time were also measured. As the results, when the voice guidance is composed that “what content” is given to the next of “place content” for the presentation order of contents it is evaluated highly in the subjective evaluations. Its voice guidance is also indicated that it is a short reaction time and a small path length of eye movement in horizontal direction when pressing the button according to the voice guidance. It means that the presentation order of contents in voice guidance influences the suitability, ease of understanding, and naturalness of Japanese language in subjective evaluation and performance of the task when pressing the button according to the voice guidance. This is important for designing the voice guidance from the practical point of view.

#### References

- [1] Japanese Industrial Standards (2007), *Ergonomics-Office work with visual display terminals (VDTs)-User guidance*, JIS Z 8523:2007.
- [2] Akatsu, H., Miki, H., Komatsubara, A. (2011), *Dynamic guidance: Synchronized auditory and visual guidance for elderly people – case study of automatic teller machines*, Ergonomics, Vol.47, No.3, 96-102 (in Japanese).
- [3] Mori, Y. and Yamaoka, T. (2008), *Experiment on interaction between elderly user and product with a card game*, Japanese Society for the Science of Design, Vol.55, No.2, 19-28 (in Japanese).