Usability Study of VUI consistent with GUI Focusing on Age-Groups

Jun Okamoto, Tomoyuki Kato, Makoto Shozakai

Information Technology Laboratory, Asahi Kasei Corporation, Atsugi, Japan

{okamoto.jd, kato.tr, shozakai.mb}@om.asahi-kasei.co.jp

Abstract

We studied the usability of a Voice User Interface (VUI) that is consistent with a Graphical User Interface (GUI), and focused on its dependency with user age-groups. Usability tests were iteratively conducted on 245 Japanese subjects with age-groups from 20s to 60s using a prototype of an in-vehicle information application. Next we calculated and analyzed statistics of the usability tests. We discuss the differences in usability with respect to age-groups and how to handle them. We propose that it is necessary to make voice guidance straightforward and to devise a VUI consistent with a GUI (VGUI) in order to let users understand the system structure. Also we found that the default design of a VGUI should be as simple as possible so that elderly users, who may be slow to learn the new system structure, are able to easily learn it.

Index Terms: voice user interface, age-group, elderly, usability, user experience, car

1. Introduction

Automatic Speech Recognition (ASR) capabilities have been widely adopted in Voice User Interfaces (VUIs) of in-vehicle information applications such as GPS-based car navigation systems. Such systems are very popular currently in Japan. However, usability issues of a VUI in such applications that focus on age-groups have not yet been discussed in detail. Berton et al. discussed a speech dialog system in Telematics and entertainment systems in cars, however, no experimental results on the usability were given [1]. Privat reported on how age affects the ASR performance, especially focusing on the elderly. However, the number of human subjects (subjects for short in the following) was quite limited and no discussion on the usability of the VUI was given [2].

Pavlovski et al. discussed user experience of a VUI and Graphical User Interface (GUI) only compared to GUI interaction [3]. They reported that natural language speech capability was the major contributor to increasing the quality of user experience in a mobile environment. Ward et al. analyzed less-studied but important topics such as root causes of timeouts and the effects of user stress in a simple dialog system [4].

We collaborated with the research project “Development of Fundamental Speech Recognition Technology” to pursue a practical usage of ASR capabilities. This project was sponsored by METI (Ministry of Economy, Trade and Industry) Japan during the years from 2006 to 2008. One of our main research goals was to establish design methodology for the universal design of a VUI consistent a GUI (VGUI for short in the following). The objective of this goal was to achieve at least a 95% task completion rate to at least 95% of users including the elderly.

A prototype of an in-vehicle information application with VGUI was developed. We then iteratively conducted usability tests for subjects with age-groups from 20s to 60s using the created prototype. Statistics of problems and the primary causes of low task completion rate were analyzed. Finally, design methodology of universal VGUIs that is easy for the elderly to use was developed.

We describe the prototype of the in-vehicle information application in Section 2. The setup of the UTs is explained in Section 3. In Section 4, some analyzed results are discussed with respect to design methodology of the universal VGUI. We examine the relationship between usability and age-group and refer to the user experience aspect. Conclusions and future works are presented in Section 5.

2. Prototype of In-Vehicle Information Application

A prototype of an in-vehicle information application was developed as a piece of PC software. Users of the prototype were able to operate it by their voice or by an 8-inch touch screen.

Implemented functions in the prototype were as follows:
- Dialing by phone number (digit string)
- Dialing by phonebook entry
- Address search
- POI(Point-of-Interest) search
- Music search

Once the talk switch is pressed, the ASR capable VGUI is activated. When a certain period of silence is detected, the ASR capability is automatically disabled. If the talk switch is pressed during playback of voice guidance, the voice guidance is immediately terminated and a new voice command is ready to be recognized. This capability is called “manual barge-in.”

The ASR engine, “VORERO” by Asahi Kasei Corporation (http://www.asahi-kasei.co.jp/vorero/en/) was used.

We conducted a preliminary study to better understand how users behave to VGUI. We asked the same questions to 24 subjects in age-groups from 20s to 50s. The subjects were asked how they react to in-vehicle information applications when they want to search POI. They responded by saying that they talk by using isolated words in most of the cases and use connected words only in cases where they cannot think of the proper isolated word. The use of natural sentences was not common.

Once the talk switch is pressed, a small window containing a microphone icon and hints on how to use VUI properly is shown on the PC display. While voice guidance is played to let users know how to operate the prototype by voice, the microphone icon is grayed-out to show that voice commands cannot be accepted. At the end of voice guidance, the microphone icon turns yellow and a beep sounds to inform users that a new voice command is ready to be accepted. If no voice is detected for ten seconds, once again the voice guidance is played to tell users how to operate it by voice. Again, if no voice commands are detected for ten seconds, the ASR capability is terminated.
3. Usability Test

3.1. Outline

Usability tests (UT in short in the following) were conducted sixteen times from August 2007 to December 2008. Three instances of these tests were done in vehicle, running on courses of a driving school to avoid traffic accidents. The rest were done in a quiet room. Nine of them were in 2007 and the rest were in 2008. In the UTs in 2008, the same tasks were repeated three times in order to observe the usability transition for subjects who were already familiar on how to operate the prototype. Therefore, the time consumed for UTs in 2008 (two hours) was nearly double compared to the time consumed for UTs in 2007.

The behaviors of subjects in all UTs were recorded by video. Furthermore, when ASR was deactivated, we asked subjects what they were thinking and what their next action was going to be. Immediately after the completion of each task we interviewed the subjects to inquire what they did and did not understand.

The prototypes were progressively modified to solve the problems we found during each UT. As a result the prototype did not look the same for each series of UTs, however, the fundamental capabilities described in Chapter 2 were not changed.

3.2. Subjects

A sample of 245 Japanese subjects with age-groups from 20s to 60s participated in the UTs. 123 were male and 122 were female. Figure 1 shows the number of subjects by each age-group. We required the subjects to meet following 4 conditions:

- They must have driver’s licenses.
- They must own now or have used in the past a GPS-based car navigation system.
- They must have mobile phones.
- They must have used PCs.

![Figure 1: Number of subjects at each age-group](image)

While users are driving the vehicle, operation through touch screen is prohibited with the exception of map operation. If users try to display a menu other than maps, a warning message is displayed on the PC screen and the display returns to the map.

Each voice command is indicated on an individual virtual button on the PC display of the prototype. For this reason, users are able to learn voice commands by touching those virtual buttons.

3.3. Prior Guidance

The prototype was explained to all subjects before the UT as prior guidance. The explanation covers implemented functions and how to activate ASR capability.

Subjects were also informed that once they press the talk switch, the voice guidance is played and the prototype is ready to accept voice commands. They were not given the following information in advance:

- They didn’t know whether they had to press the talk switch every time they talk or just once at the very beginning of each task.
- They didn’t know that the words indicated on the virtual buttons in the display correspond to voice commands.
- They didn’t know of the manual barge-in function, which means that the voice guidance can be terminated when they press the talk switch.

In order to observe the usability of the VGUI, we had to let the subjects know that only operation accessible by voice was enabled and operation by touching screen was disabled.

3.4. Tasks

Tasks given in the UTs are listed below:

- Dialing a phone number with the GUI (Dialing number w/ GUI for short in the following)
- Dialing by phonebook with the GUI (Dialing phonebook w/ GUI for short in the following)
- Dialing by phone number without the GUI (Dialing number w/o GUI for short in the following)
- Dialing by phonebook without the GUI (Dialing phonebook w/o GUI for short in the following)
- Address search
- POI search
- Music search

We assumed that tasks w/o GUI corresponded to hands-busy and eyes-busy situations of driving the vehicle. The task requirement was presented to subjects by a task instruction sheet where the present situation and what to do next was described. All tasks were presented to all subjects in the same order to eliminate any effect caused by task order.

No time limitation for task completion was set. If subjects thought the required task was completed, they were asked to declare orally that it has been completed. If they thought that they could not complete the task without help, they were asked to declare orally that they had given up.

3.5. Problem importance level

Problems that were discovered in the UTs were summarized and then classified depending on their root causes. Finally, they were categorized into three different levels depending on their importance: light, medium and serious. Definitions of problems at these three importance levels are described in Table 1.

4. Discussion

The total number of observed problems in the series of UTs was 2,173.
Table 1. Definition of problem importance levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>Observed problem had little effect on task completion</td>
</tr>
<tr>
<td>Medium</td>
<td>Observed problem in a task that was completed without any assistance but required a large consumption of time</td>
</tr>
<tr>
<td>Serious</td>
<td>Observed problem in a task which was incomplete or where the subject misunderstood the task</td>
</tr>
</tbody>
</table>

Figure 2: Ratio of problems at the three importance levels

Figure 3: Average occurrence of problems per age-group

Figure 4: Number of problems per task

Figure 5: Tendency of serious problems by age-group

Figure 2 shows the ratio of problems at three importance levels. Two thirds (2/3) of the problems belonged to medium importance level.

The average number of problems that occurred per subject by age-group is shown in Figure 3. There is no significant difference between females and males. On the other hand, the average number of problems that occurred per each subject increased as the age-group became higher.

Figure 4 shows the number of problems per each task. “Common” in Figure 4 means common problems shared in multiple tasks, such as unsynchronized timing between the prototype’s operation mode and subject’s voice commands. They often talked to the prototype when it was not ready to accept the voice commands for ASR capability.

Figure 4 suggests that complicated tasks such as address search or POI search gave more problems than simpler tasks. Furthermore, it is interesting to note that the number of problems in the case of dialing by phone number w/ GUI was much larger than that of w/o GUI. Information that the subject can obtain w/o GUI was less than that w/ GUI, however, the results showed the opposite. This suggested that the order of the tasks strongly affected the tasks’ completion. Throughout the interview that we had with subjects, it turned out that they became accustomed to VGUI structure and its operation flow. In these UTs, subjects were only informed by prior guidance that once they press the talk switch the voice guidance is played and the prototype is ready to accept voice commands. Therefore, at the beginning they tried to behave based on their experience and the mental model they acquired when they used GPS-based car navigation systems in the past. However, as the mental model didn’t match with the prototype, they operated the VGUI in an unsuitable way. Because dialing by phone number w/ GUI was the first task for most of the subjects, they experienced more problems in this task.

Figure 5 shows the tendency of the serious problems by age-group. It reveals that 50% of serious problems in the dialing number w/ GUI are given by the 60s age-group. On the contrary, the 20s gave no serious problems. This suggests that speed of learning the new system structure differs depending on age-group.

Causes of serious problems are classified in Figure 6. It suggests that there are many causes related to voice guidance and structure understanding. Subjects often misunderstood what voice guidance meant because it lacked a straightforward representation. It is necessary to explain how to operate the system by showing straightforward examples. With regard to structure, simplifying the system leads to a better understanding for subjects. However, there are some cases where it is difficult to simplify the system. Therefore it is
subject whose task completion rate was 85% while the rest of the subjects were over 95%. Thus, we confirmed that the goal of our research (at least 95% task completion rate for at least 95% users) was accomplished. Additionally, we asked the subjects to provide a subjective score with regard to their interest to continue using the in-vehicle information appliance by voice on a scale of 1 (They don’t want to keep using it) to 5 (They want to keep using it). We asked this question both before the UT as an expectation and after the UT as an experience. The result is shown in Figure 8. While the percentage of subjects who showed interest with an objective score of 4 or 5 was 64% as an expectation, it was increased to 79% as an experience. This encouraging result suggests that the established design methodology helped increase user experience. However, the primary complaint given by subjects in their 60s after the UT was still with respect to voice guidance. They reported that it was straightforward but long. Once they learned how to use the VGUI, they didn’t need the repetition of voice guidance any more.

5. Conclusions

We analyzed problems and root causes observed in UTs of 245 Japanese subjects using an in-vehicle information application. Results showed that it is necessary to make voice guidance straightforward to assist understanding the structure of the VGUI system and to devise the system in order to allow subjects to understand the structure easily. Also, we found that the design of the default structure of a VGUI should be as simple as possible so that elderly users, who may be slow to learn the new system structure, are able to easily learn it.

On the other hand, it was interesting to find that experienced users, who were familiar with the system, felt irritated about using the VGUI that was devised for easier understanding of the system structure. It is obvious that users are soon able to learn the structure of the VGUI if it is designed by the proposed methodology. If the VGUI stays at the easier level, the experienced users will feel unsatisfied. It suggests that VGUI should change its behavior by dynamically judging the user’s proficiency on how to use the VGUI. To pursue the user experience of a VGUI, another research effort is necessary. How to measure the user’s proficiency iteratively will be one of the most interesting research topics in these future efforts.

6. Acknowledgements

This research was supported by the METI: “Development of Fundamental Speech Recognition Technology,” and was supervised by Prof. Furui at Tokyo Institute of Technology and Prof. Kobayashi at Waseda University.

7. References


