Evaluating Habituation Effect on Conversational Voice Control in Home Network System

Noriyuki MATSUBARA†, Shinsuke MATSUMOTO†, and Masahide NAKAMURA†

† Kobe University Rokkoudai 1–1, Nada, Kobe, Hyogo, 657–8501 Japan

Abstract The voice control is an easy and effective method to operate heterogeneous devices in the home network system (HNS). However, as the user is getting used to the interface, too much feedback from the system annoy the user and decline the usability. In this paper, we empirically evaluate the habituation effect on the voice control in the HNS. Specifically, we characterize the habituation effect by the time since the system begins the feedback until the user issues the next command. Based on the empirical study, we consider a voice feedback method that adapts the individual habituation.

Key words Home Network System, remote controller, voice interface, habituation effect

1. INTRODUCTION

The Home Network System (HNS) has been studied extensively as the next-generation ubiquitous application. The HNS connects various home appliances and sensors to the network, and provides comfortable services [1].

One of the important research topics in the HNS is the user interface (UI). The conventional remote controllers and the integrated control panels have limitations in operating a number of heterogeneous appliances and services [2]. The *voice interface* is a promising UI for the HNS, which is being focused recently. Several methods to operate home appliances via voice have been proposed [3]. Basically, the voice interface for the HNS can be implemented by associating certain words spoken by a user (i.e., *voice command*) to an API of the HNS that invokes an appliance operation. The voice interface can be adapted easily to various configurations of the HNS. This is because the behavior of the voice interface is determined by the association of a voice command and an API. Also, various users can learn it easily, since various commands can be issued by the speech only.

In our previous work [4], we have developed the voice interface for our practical HNS, called CS27-HNS. The method employs *mixed-initiative interaction* [5]. The proposed method constructs voice commands through the interaction with a user. So it is unnecessary for the user to memorize a lot of voice commands. Also, from user speech like “it’s hot” and “it’s dark”, the system automatically suggests user’s implicit requirements such as “I want the fan or air-conditioning”, or “I want the light to be turned on”, and recommends the commands.

The advantage of the voice interface with mixed-initiative interaction is that any non-expert user can operate appliances easily. However, for those who got used to the voice interface, it is not always comfortable to use. Every time a user operates an appliance, the system speaks the feedback. The feedback occurs whenever the user selects an appliance, an operation, and parameters. As a result, the interaction may annoy the user, because the system spends several tens of seconds to finish an appliance operation.

To achieve high usability for the user who got used to the interface, we have to introduce a method that detects *user habituation* and optimizes the system feedback.

In this paper, we empirically evaluate the habituation effects of the voice interface in the HNS. In the experiment, we instructed subjects to repeat using the proposed voice interface in the CS27-HNS [6]. We then observe the change of time to spend to interact with the system. Through the analysis of the habituation to the system, we reveal how the habituation occurs and can be detected. Furthermore, we consider methods to improve the usability the users who get used to the voice interface.

2. PRELIMINARIES

2.1 Home Network System (HNS)

The HNS consists of *networked home appliances* such as fans, air-conditioners, lights, curtains, TVs. It also has *networked sensors* including thermometers and illuminometers. A home server manages all the networked devices in the HNS. Every appliance exhibits the application program interface (API) to the network. The API allows users and external agents to monitor and control the appliance. In our CS27-

HNS, various appliances including a fan, an air-conditioning, a light, a curtain and a TV exhibits the API as Web services [6]. So we can operate them through various platforms.

2. 2 Conventional Voice Interface for Appliances

The voice interface has been attracting attention as one of the next-generation UI, since it doesn't require the user any special device. It recognizes the user speech using voice recognition engine. An appliance operation is triggered by the association of the recognition results and the API invocation. Therefore, this interface can be introduced to different environments of the HNS with relative ease.

For example the voice interface for recording videos [3], the home appliance operation interface for quadriplegia by direction of head and voice [7], etc are proposed as the interface for home appliances.

2. 3 Previous Work [4]

We have developed the voice interface with mixed-initiative interaction in the HNS in the previous work [4]. The merit of the voice interface with mixed-initiative interaction is that the users don't need to learn prior knowledge of voice commands. The voice commands are built through interactions between the user and the system. By this method of interaction, even the users who have never used the voice interface can operate appliances to only speak an appliance name and an operation name which the system feeds back.

The following shows a typical workflow of the proposed voice interface.

(1) The system feeds back the list of available home appliance names by speech.

(2) The user speaks a name of home appliance which he want to operate.

(3) The system feeds back the list of operation names of the home appliance by speech.

(4) The user speaks an operation name.

(5) The system invokes the appliance operation.

In addition to the above explicit command operations, the proposed interface identifies user's implicit requirements from conversations of users. The conversation include dissatisfaction with the present circumstances and expectation to the ideal state, for example, "It's hot", "It's dark" , etc.

Also, to achieve high usability, the proposed interface enables asynchronous interaction. The system can accept user speech even while the system is speaking the feedback. Therefore, before the user listens to the feedback to the end, the user can speak the next voice command.

2. 4 Problem

A problem of the proposed interface is the decline of the usability by the habituation. The advantage of the proposed interface is that the user don't need to learn voice commands. However, it is not always comfortable for the users who got used to the voice interface.

When a user uses the proposed interface, the system spends time to issue the feedback. The voice interaction between the user and the system occurs in every step of selections of an appliance, an operation, and parameters. The proposed interface adopts asynchronous interaction. However, the system needs at least 2 steps to invoke one appliance operation. Especially, as for the routine operations that the user invokes every day, the habituated user must be bored with the voice interaction itself.

To achieve high usability for the habituated users, we need to introduce a method that detects user habituation automatically, and changes the system behaviors.

3. PROPOSED METHOD

3. 1 Key Idea

To cope with the problem in Section 2. 4, we aim to detect the habituation effect in the appliance operations by the voice interface in this paper.

The voice interface with mixed-initiative interaction can be used easily by various users. This is because the user only speaks the next command by relying on the information provided as the voice feedback from the system. However, as the user repeats the appliance operations, he learns the feedback from the voice interface. Eventually, he gets used to appliance operations and he does not need the careful feedbacks. If the interface can detect the habituation of the user, it can reduce the wasteful long feedback to optimize the interaction overhead. Feeding back the voice commands, which the user has learned, from the voice interface is very annoying for the user who habituate himself to appliance operations by the voice interface. Therefore, to adjust feedback from the voice interface is important from this perspective that the interface reduces the stress of the users.

We explain the details of this key idea in Section 3. 2.

3. 2 Detecting Habituation to Appliance Operations by Voice Interface

Generally, a user tends to get used to appliance operations by the voice interface as he repeats them. The user who has habituated himself to appliance operations speaks the next voice for the feedback from the voice interface at early timing. Conversely, the user who hasn't already got used to appliance operations at all speaks it after he listens the feedback from the voice interface firmly until it finishes. Therefore, a promising metric to detect the habituation is the time interval from one voice command to the next command. As the user is getting used to the interface, the time interval is expected to become shorter.

At this point, we introduce three kinds of time spent within the voice interface with mixed-initiative interaction.

Time of user speech(T_{user_speech}) : Time from the user starts speaking a voice command till he finishes speaking. T_{user_speech} may vary depending on the length of the voice commands, but the variation is not so significant.

Time of system process($T_{sys\_process}$) : Time spent by the system to recognize user voice command, synthesize the voice feedback and invoke an appliance operation. $T_{sys\_process}$ has no significant variation, since it depends on the system.

Time of system speech(T_{sys_speech}) : Time from the system starts speaking a feedback until the system accepts the next voice command from the user.

For every voice interaction, the above three types of time (intervals) appear in the order of $T_{user\_speech} \rightarrow T_{sys\_process} \rightarrow T_{sys\_speech}$.

We also suppose that **the number of mistakes(N_{miss})** and **the number of auxiliary commands(N_{aux})** will be reduced as the user is getting used to the interface. The auxiliary commands refer to commands to start/stop the voice interface, command to show help, etc.

We propose to use two metrics to evaluate the habituation effects. The one is **the sum of** T_{sys_speech} within a complete sequence of voice interactions. Another is **the total number of** N_{miss} **and** N_{aux} within a complete sequence of voice interactions.

T

3.3 Implementation

We implemented the voice interface which logs appliance operations and times T_{sys_speech} on the basis of the proposed key idea. We have the following the implementation environment of the proposed system.

The voice recognition engine : We used Julius4.1.2 [8] as the voice recognition engine. The feature of Julius is that we can define freely grammar and vocabulary which we want to recognize. In this paper, we registered 7 grammars and 52 words.

The voice synthesis engine : We used VoiceText Engine manufactured by PENTAX as the voice synthesis engine which outputs voice feedback.

The language for implementation : We implemented the application to use the voice recognition engine in Java language on Eclipse.

3.4 Steps and Workflow of Voice Interactions

We show 10 steps of implemented voice interactions in Table 1. These steps are executed according to the workflow presented in Figure 1.

At each step, the interface measures T_{sys_speech}, and accumulates the sum of T_{sys_speech}.

Table 2 Appliance operations of each task

appliance	Task 1	Task 2
curtain	Open	Open
fan	Power on Maximize air volume –	Power on Maximize air volume Swing mode
TV	–	Power on Switch input mode to PC
air-conditioner	–	Turn on cooling mode

4. EVALUATION EXPERIMENT

4.1 Overview

To evaluate empirically the habituation effect, we have conducted an experiment, where subjects operate home appliances connected to the HNS using the proposed system. In the experiment, the subjects performed practice at first, where each subject turns on a light. Next, we asked them to perform two kinds of tasks : Task 1 and Task 2. In Task 1, every subject performed appliance operations as instructed. In Task 2, each subject freely operated the appliances to realize the designated environment. Table 2 summarizes the concrete appliance operations in the tasks. We instructed the subject to perform the practice once, and each of Task 1 and Task 2 for five times. Also, we reset all the appliances after one trial was finished. The total 7 subjects participated in the experiment. All of them were 20's years old, and none of them were familiar with the voice interface.

We have collected following metrics as evaluation criteria to evaluate the habituation effect.

Sum of T_{sys_speech}**(**$M_{sys\_speech}(i)$**) :** The sum of T_{sys_speech} which the subject spent to finish i-th trial of a task.

Total number of N_{miss} **and** N_{aux}**(**M_{miss}**) :** Total number of operation mistakes and auxiliary commands until the subject finishes one trial.

We have timed them from the subjects speak "*OnseiComputer*" at the step of Starting the interface to the step of Stopping the interface.

4.2 Experiment Environment

We used the CS27-HNS in the experiment. The appliances used in Task 1 were a curtain and a fan. The appliances used in Task 2 were the curtain, the fan, a TV and an air-conditioner. We used a Bluetooth headset Voyager510 manufactured by Plantronics as the voice input device. The headset provides hands-free microphone operation via wireless connection.

4.3 Procedure of Experiment

We describe the procedure of the experiment as follows.

1. We explained the background of the experiment (the HNS etc) to subjects, and obtained their informed consents.

Name of step		Contents of step
Starting the interface		The step that the user speaks the auxiliary command *"Onsei Computer"* and starts the interface.
Registering the user name		The next step of Starting the interface. The user registers his name to log his appliance operations.
Operate an appliance	Selecting an appliance	The next step of Registering the user name. The user inputs an appliance command.
	Selecting an operation of the appliance	The next step of Selecting an appliance. The user inputs an operation command.
	Selecting a parameter of the operation	The next step of Selecting an operation of the appliance. The user inputs a parameter command.
	Selecting an appliance and an operation	The next step of Registering the user name. The user inputs an appliance command and an operation command simultaneously.
	Selecting an appliance, an operation and a parameter	The next step of Registering the user name. The user inputs an appliance command, an operation command and a parameter command simultaneously.
	Invoking the appliance operation	The step that the user completes the necessary commands and invokes the appliance operation.
	Mistaking operations or Inputting an auxiliary command	The step that the user mistakes operations or speaks an auxiliary command, *"Onsei Computer"*, *"Sosa Hoho"*, *"Cancel"*.
Stopping the interface		The step that the user speaks the auxiliary command *"Onsei Computer"* and stops the interface.

Table 1 Steps of voice interactions with the voice interface

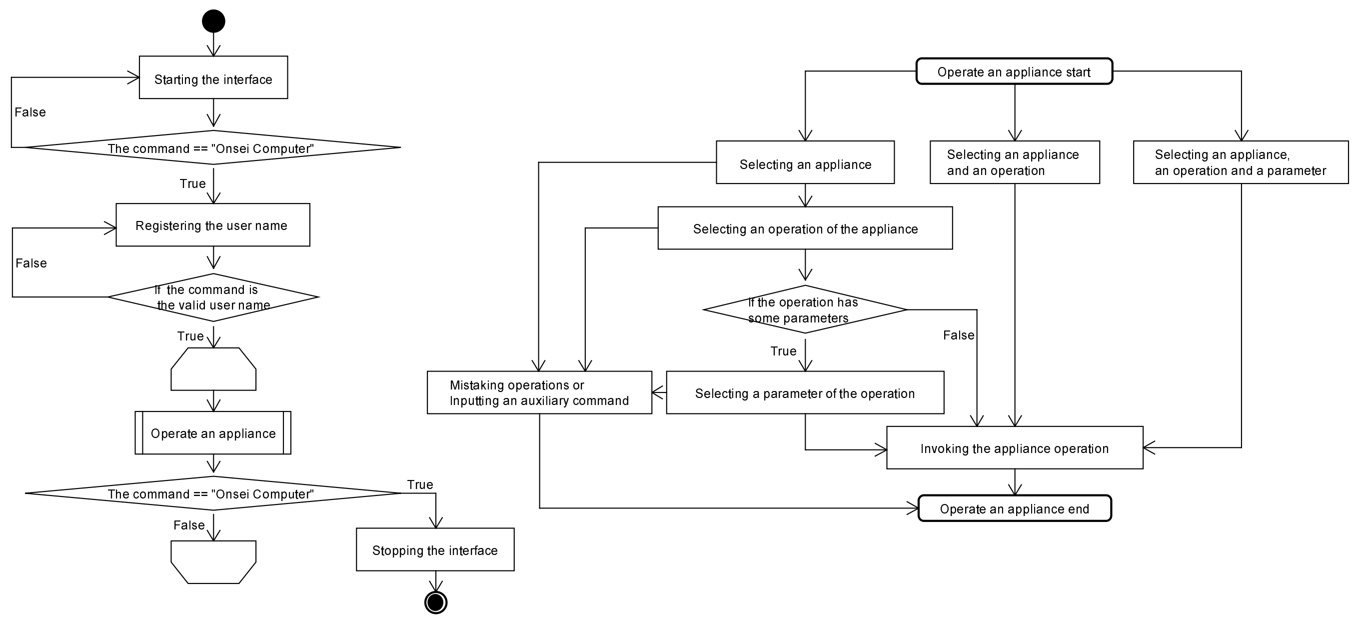


Figure 1 Workflow of voice interactions with the voice interface

2. We explained the basic operation procedures of the proposed interface.

3. We gave instructions of the experiment and the tasks.

4. The subjects performed preliminary practice.

5. The subjects performed Task 1 with the voice interface.

6. The subjects performed Task 2 with the voice interface.

The explanation at Step 2 is as follows.

- At first, if the subjects speak only a home appliance name, they can perform basic operations.
- If the subjects speak an auxiliary command "*OnseiComputer*", they can start and stop the voice interface.
- If the subjects speak an auxiliary command "*SosaHoho*", the system will feed back the basic operation method.
- If the subjects speak an auxiliary command "*KadenPickUp*", the system will feed back the list of available home appliances.

4.4 Experimental Results

Figure 2 shows a graph plotting M_{sys_speech} taken by each user to perform 5 trials of Task 1. Also, Figure 3 plots M_{miss} of each 7 subjects in Task 1. In the graphs, A, B, ... ,G represent the 7 subjects. Similarly, Figures 4 and 5 are those measured in Task 2.

We first explain Task 1. If we focus M_{sys_speech} in Figure 2, we can see that the value of all subjects is reduced, as they repeat the trials. Also, if we focus M_{miss} in Figure 3, we can see that the value of 5 subjects except D and G converges in 0 at the final 5th trial.

Next, we explain Task 2. If we focus M_{sys_speech} in Figure 4, we can see that the value of all subjects becomes the least at the 5th trial. Also, if we focus M_{miss} in Figure 5, we can see that the value of all subjects converges in 0 at the final 5th trial.

Thus, the more the subjects repeat the trials, the smaller the value of M_{sys_speech} and M_{miss} in both of Task 1 and Task 2 become. Therefore, we can observe the habituation effects in these metrics.

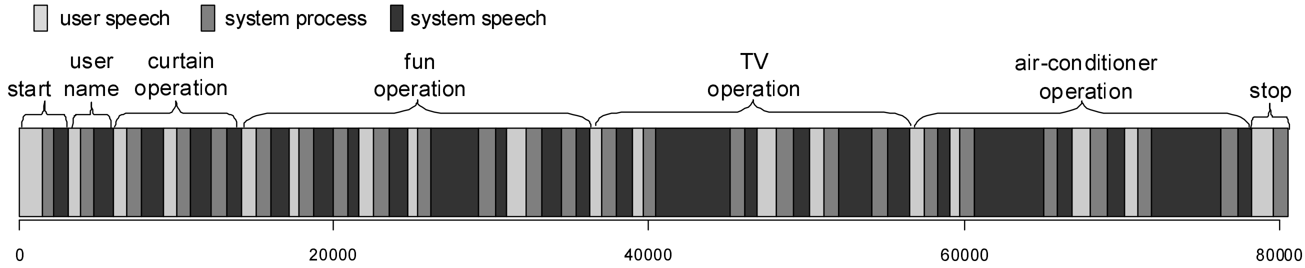


Figure 6 Speech process at 5th trial in Task 2 of subject B

Figure 2 Sum of T_{sys_speech} in Task 1(M_{sys_speech})

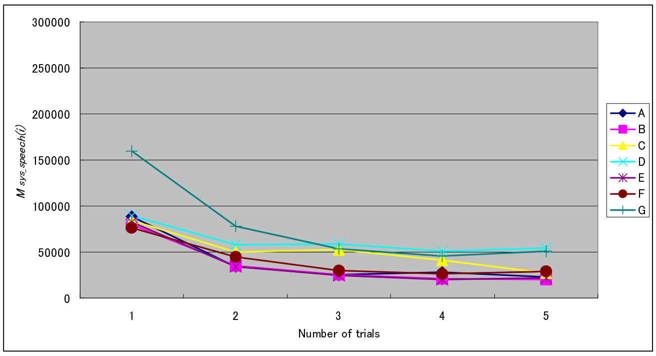


Figure 3 Total number of N_{miss} and N_{aux} in Task 1(M_{miss})

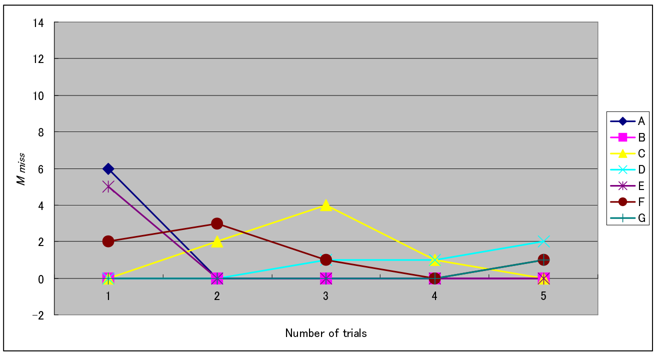
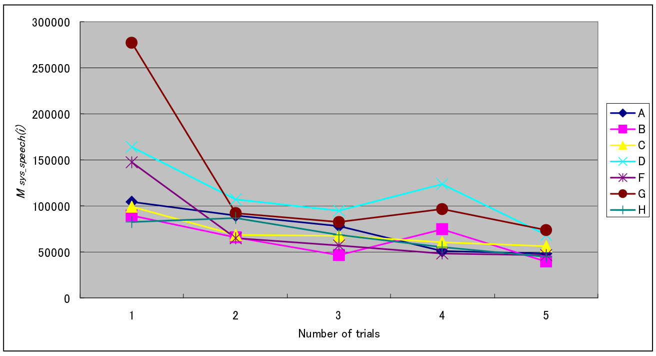


Figure 4 Sum of T_{sys_speech} in Task 2(M_{sys_speech})



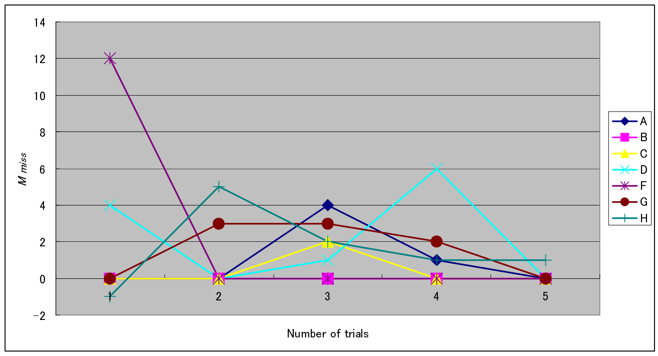
5. Discussion

5.1 Essential Factors Characterizing Habituation

5.1.1 Time Factor

If we focus Figures 2 and 4, M_{sys_speech} of all subjects was reduced as they repeated the trial. Although the individual differences exist, M_{sys_speech} can well characterize the habituation effects. We expected that M_{sys_speech} in Task 2 would be larger than that in Task 1. However, there was no significant differences between Figure 4 and Figure 2. We consider that this is because the habituation obtained in Task 1 was taken over to the Task 2.

Figure 5 Total number of N_{miss} and N_{aux} in Task 2(M_{miss})



Thus, there seems to be two kinds of habituation. One is "habituation in the task", which is obtained by the repeat of the same routines. Another is "habituation between tasks", which is globally obtained by performing various tasks.

5.1.2 Reliability Factor

On the other hand, M_{miss} in Figure 3 and Figure 5 tend to converge scarcely. When we analyze it, the subjects are prone to different speech mistake between in the first half and in the latter half. There are large "Operation mistakes by system misrecognition" in the first half. And there are large "Speech mistakes by user habituation" in the latter half.

At first, "Operation mistakes by system misrecognition" in the first half occur when the system misrecognize user speech. As the subjects repeat experiments, it had resolved by user intention to speak clearly and speak a command, which is easy to misrecognized for example "on/off", crisply.

On the other hand, "Speech mistakes by user habituation" in the latter half occur when the speech timing of the subjects are too early. The process flows of the proposed interface are 1.user speech, 2.system recognition and process of user speech and 3.system speech. During 3.system speech, the system can accept user speech However, during 2.system recognition and process of user speech, it can't. Therefore, we think that the system needs the improvements that the processes are arranged in a line or we optimize the faster process of speech recognition.

5. 1. 3 Limitations of Interactive Voice Interface

To investigate the breakdown of the time to spend whole appliance operations, we show speech process at 5th trial in Task 2 of subject B who have finished it fastest in Figure 6. It spent about 80 seconds to finish the task overall. Especially, it spent much time to operate the appliances which have many operations, for example the TV, the air-conditioner, etc. Also whole time of 20% is user speech and it of 30% is system recognition and process of user speech and it of 50% is system speech. So, the time to spent system speech and system recognition and process cause the decline of usability. We think that the decline of usability by this habituation is the limitations of the voice interface with mixed-initiative interaction.

5. 2 Detecting Habituation

To consider how to detect the habituation automatically by the proposed interface, we interviewed the subjects about how many times they felt the habituation in each task. As a result, subjects A to G respectively felt the habituation at 3rd, 2nd, 2nd, 3rd, 2nd, 4th, 3rd trials in Task 1. As for in Task 2, they felt the habituation at 3rd, 3rd, 2nd, 5th, 3rd, 4th, 3rd trials, respectively. Suppose that a subject felt the habituation at k-th trial. Then a difference $H_{time}(k) = M_{sys\_speech}(k-1)$ - $M_{sys\_speech}(k)$ would be a good metric to capture the habituation. Based on the answer, the average of $H_{time}(k)$ was about 5.6 seconds. Therefore, we can empirically detect the habituation when $H_{time}(i)$ is close to 5.6 seconds for arbitrary i. In this way, we think that we can detect the user habituation to the interface by recording the convergence degree of T_{sys_speech}.

5. 3 Feedback for Habituation

To control the decline of usability by the habituation, T_{sys_speech} need to be held down. Concretely, we propose how to discontinue the feedback of operation commands or parameter commands which possible to designate in the step which the habituation is detected. Also, to reduce the number of interaction, we propose how to speak an appliance command, an operation command and a parameter command simultaneously. The simultaneous speech has been implemented in the proposed interface. However, no subject noticed the existence of it. We think that the system should urge the user to speak the commands simultaneously when the habituation is detected.

Also, we propose how to change the feedback procedure of operation commands or parameter commands by user preference and frequency of use.

6. CONCLUSION

In this paper, we have showed a problem of the proposed interface in Section 2. 4. We have proposed a key idea K in Section 3. 1, and a method realizing K to resolve the problem in Section 3. 2. Also, we have experimented to evaluate the proposed method in Chapter 4., and discussed the experimental results in Section 4. 4 in Chapter 5..

Our future work includes the setting of the specific threshold value to detect the user habituation to appliance operations by the voice interface, the implementation of the method to deal with the habituation, the control of the timing of user speech and the decrease in the incorrect operations.

The voice interface to operate the home appliances hasn't been generalized yet. However, we think that it would be used routinely by more people if we resolved these problems.

ACKNOWLEDGMENTS This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research(B) (No.23300009), Young Scientists (B) (No.21700077), Research Activity Start-up (No.22800042)], and Hyogo Science and Technology Association.

REFERENCES

[1] H. Morikawa, "Ubiquitous network and the role of wireless," The Journal of the Institute of Electronics, Information, and Communication Engineers, vol.87, no.5, pp.356–361, 2004.

[2] The Accessible Design Foundation of Japan, "Inconvenience investigation database," `http://www.kyoyohin.org/02_syougai/index.php`, 03 2008.

[3] Y. Watanebe, Y. Sekiguchi, and Y. Suzuki, "On the speech dialogue function of household electric appliances goods as an example of video tape recorder," Transactions of Information Processing Society of Japan, vol.44, no.11, pp.2690–2698, 2003.

[4] N. Matsubara, K. Egami, H. Igaki, and M. Nakamura, "Interactive voice interface for eliciting and estimating implicit user requirements in home network system," TECHNICAL REPORT OF IEICE, vol.109, pp.061–066, 2010.

[5] T. Kawahara and M. Araki, The voice interaction system, Ohmsha, 2006.

[6] M. Nakamura, A. Tanaka, H. Igaki, H. Tamada, and K. Matsumoto, "Constructing home network systems and integrated services using legacy home appliances and web services," International Journal of Web Services Research, vol.5, no.1, pp.82–98, Jan. 2008.

[7] H. Ito, "Home appliance controller for quadriplegia by direction of head and voice," Journal of Human Interface Society, vol.4, no.1, pp.25–31, 2002.

[8] T. Kawahara and A. Lee, "Open-source speech recognition software julius," Journal of Japanese Society for Artificial Intelligence, vol.20, no.1, pp.41–49, 2005.