Effects of Auditory Robot Interfaces on Human-Robot Interaction

Comparing Speech Interface with Non-speech Audio for Korean Elderly Subjects

Sonya S. Kwak(sonakwak@kaist.ac.kr), Myung Suk Kim(kimmyungsuk@kaist.ac.kr).
Dept. of Industrial Design, Korea Advanced Institute of Science and Technology (KAIST), KOREA

Abstract : With the advancement of robotics and the ageing society, the necessity of tool-like service robots for elders is emerging. For service robots which appear in human's daily life, human-robot interaction(HRI) is essential factor to design. Particularly, auditory interface can be an effective interaction way for elders who are not accustomed to new device. The objectives of this study is to find whether speech interface or non-speech audio are more preferred by users and support more effective task performance. Two tasks with corresponding speech and non-speech audio were given to participants for experiments. The correctness and the time spent to complete each task were measured, and questionnaires about preferences and task performances were asked. Participants exhibited more preferences on speech interface than non-speech audio and completed tasks in shorter time mostly with speech interface.

Keyword : Tool-like Service Robots, Speech Interface, Non-speech Audio

1. Introduction

Due to the development of robotics and growing numbers of aging people[1,2], the necessity of service robots which assist silver generation's daily life is increasing. Among various types of service robots, tool-like robots, which operate as "smart appliances"[3] are effective form of service robots having similar shape of existing products with autonomy. Because service robots appear human's daily life for home monitoring and security, attractive and understandable interaction design is essential[1] to service robots. Especially, because silver generation become weak and unfamiliar with newly developed technology or device[4], auditory interaction can be useful and efficient interaction way for silver generation.

The purpose of this study is to find more ideal human-robot interaction by comparing speech interface and non-speech audio.

2. Related Works- Speech Interface vs. Non-speech Audio

Auditory interface is emerging as an efficient interaction way for ubiquitous environment or mobile devices[5,6]. Auditory interface can be divided into speech interface and non-speech audio. Speech interface is efficient when transmitting textual information; while it is often improper to be used in public environments; and designers should avoid any output word which is used for input[7]. Non-speech audio is efficient for speedy and urgent transmission or continuous monitoring of background information[7], and it provides unobtrusive feedback[8]. In sum, speech interface is more efficient for conveying text-based information while non-speech audio is more effective as navigation cues[9].

3. Experiment

Based on the study, "Assistive Robotics and an Ecology of Elders Living in Their Homes"[1], researchers of 'Project on People and Robots' at Carnegie Mellon University(CMU) developed the 'Sense Chair'[10] which is a tool-like robot to assist elders living. As a cooperative research with CMU, this study investigates Korean elders' preferences and task performances when using speech interface and non-speech audio for tool-like service robots.

3.1. Hypothesis

Speech interface will show higher user preferences as well as task performances.

3.2. Method

(1) Participants

Approximately eight participants who use Korean as a native language and aged over 65 were recruited from 'Youseoung-Gu welfare center for elders'. The proportion of males and females were the same, and elders whose auditory senses were extremely debilitated were excluded.

(2) Procedure

The experiment is composed of independent variables(speech interface vs. non-speech audio) and dependent variables(preferences and task performances). The experiment setting is established in the welfare center for elders. A lounge chair which includes speakers inside for playing recorded sound, a portable stove, and a television schedule section of newspaper are prepared for tasks. In this study participants will complete two tasks: turning off a boiling kettle and finding a TV listing. Two tasks were given randomly according to speech interface and non-speech audio.

Detailed procedure is as follows:

1) Functions of the chair robot and the objectives of the study are explained.

2) A preview of the experiment(alarming a visitor at the door) is illustrated.

Table 1. Speech Interfaces and Non-Speech Audios in Task 1 and 2

Task1: Controlling Appliances (Turning Off the Stove)	
Speech Interface	"The water is boiling! You'd better turn off the stove"
Non-Speech Audio	Boiling water / Piercing whistle of a kettle boiling
Task2: Finding a TV listing	
Speech Interface	"Can you find the listing for 'Lovely Darling'?"
Non-Speech Audio	'Lovely Darling' Theme song.

3) In random order, scenarios of task1 or 2 are explained, and corresponding speech or non-speech sound are played. The correctness and the time consumed to complete each task are measured.

Speech interfaces and non-speech audios used in task1 and

2 are shown in [table 3-1].
4) After a participant complete both tasks, four auditory interfaces are replayed in the same order of sounds played in the previous experiment, and survey about each sound is executed. The survey is composed of questions related to user preferences and task performances, and questions about personal information.

(3) Measures

1) Measure of Task Performances

In order to measure task performances, measurement of participants' correctness and time to complete each task and questionnaires were used. Regarding correctness, two chances were given. If participants could not complete the task twice, it is regarded as failure. 'Ease of use(5 questions)' and 'Credibility(4)' were used in seven-scale based questionnaires.

2) Measure of User Preferences

In order to measure user preferences, seven-scale based questionnaires were used. The questionnaires are composed of 'preferences(5)', 'social(5)', 'intelligent(4)', and 'potency(3)'.

3.3. Analysis

(1) Analysis of Task Performances

Results of the correctness of each task and the time spent to complete each task are indicated in [table 3-2]. All participants display success in task1 while five participants result in failure in task2 with non-speech audio. In task1, participants 1, 4, 6, 7, 8 take shorter time with speech interface while participants 2, 3, 5 take shorter time with non-speech audio. In task2, all participants display higher task performances with speech interface taking shorter time in completing the task.

Table 2. Correctness and Time Spent to Complete Each Task

(sec)	Task 1		Task 2	
	Speech	Non-speech	Speech	Non-speech
Sub1	1.89	2.48	4.21	Failure
Sub2	3.33	3.24	8.71	37.26
Sub3	3.90	3.84	2.49	Failure
Sub4	2.86	4.08	Failure(1st), 4.12(2nd)	Failure
Sub5	3.30	3.09	4.14	16.68
Sub6	3.31	18.78	4.81	34.14
Sub7	3.70	3.80	6.16	Failure
Sub8	2.93	10.28	9.52	Failure
Ave	3.15	6.20	6.83	29.36

Table 3. Task Performances Measured by Questionnaires

(pts.) total 7pts.	Task 1		Task 2	
	Speech	Non-speech	Speech	Non-speech
Sub1	6.43	5.83	6.65	3.75
Sub2	6.63	5.58	6.68	1.55
Sub3	7.00	7.00	6.90	5.00
Sub4	6.40	5.70	5.28	2.13
Sub5	6.63	5.70	6.50	4.43
Sub6	6.88	5.73	7.00	4.95
Sub7	7.00	6.20	6.15	2.75
Sub8	6.80	6.80	6.53	2.10
Ave	6.72	6.07	6.46	3.33

Results of the task performances through questionnaires are shown in [table 3-3]. Except participant 3 and 8 who score speech interface and non-speech audio in task1 with equal points, all the other participants give higher marks on speech interface in both task1 and 2. Regarding the difference between average of speech interface and non-speech audio, the difference in task1(0.65) is smaller than that in task2(3.13). This indicates that task performance is not only influenced by speech vs. non-speech but also affected by types of tasks. About the total average grades, higher task performances are shown in the order of speech interface in task1(6.72), speech interface in task2(6.46), non-speech audio in task1(6.07), and non-speech audio in task2(3.33).

(2) Analysis of User Preferences

Results of user preferences through questionnaires are displayed in [table 3-4]. In all cases, all participants prefer speech interface to non-speech audio. About the difference between average of speech interface and non-speech audio, the difference in task1(1.27) is smaller than that in task2(1.31). This indicates that user preferences are affected by types of tasks as well as speech vs. non-speech. About the total average according to types of tasks and auditory interfaces, higher preferences are exhibited in the order of speech interface in task1(6.08), speech interface in task2(5.68), non-speech audio in task1(4.81), and non-speech audio in task2(4.37). About the total average according to questionnaire items, higher grades are displayed in the order of intelligent(5.38)>social(5.20)>preferences(5.12)>potency(4.68). This result implies that user preferences are more influenced by questionnaire items with higher grades.

Table 4. User Preferences Measured by Questionnaires

(pts.)	Task 1		Task 2	
	Speech	Non-speech	Speech	Non-speech
Sub1	6.09	5.36	6.23	4.86
Sub2	6.42	4.67	6.31	2.56
Sub3	6.83	5.50	5.93	5.18
Sub4	5.65	4.90	5.41	4.78
Sub5	6.52	4.10	5.63	4.36
Sub6	6.00	3.87	5.36	4.77
Sub7	5.47	5.15	5.14	3.69
Sub8	5.65	4.90	5.41	4.78
Ave	6.08	4.81	5.68	4.37

4. Conclusion and Further Study

In this study, speech interface display higher task performances and user preferences than non-speech audio, and task1 induces higher task performances and user preferences than task2. These results indicate that types of tasks as well as auditory interfaces affect task performances and user preferences. As further study, extended experiments in more various types of tasks can be executed, and experiment results with Korean participants and those with American participants can be compared.

References

[1] Forlizzi, J., DiSalvo, C., Gemperle, F., Assistive Robotics and an Ecology of Elders Living in Their Homes, Journal of Human Computer Interaction 16 (2004).
[2] http://100.naver.com/100.php?id=13176
[3] Fong, T., Notbakhsh, I., Dautenhahn, K., A Survey of Socially Interactive Robots, Robotics and Autonomous Systems 42 (2003).
[4] Hirsch T. et all. The ELDer Project: Social, Emotional, and Environmental Factors in the Design of Eldercare, in: Proceedings of Conferences on Universal Usability, 2000.
[5] Rosenfeld, R., Olsen, D., and Rudnicky, A., Universal Speech Interfaces, Interactions, November + December, 2001.
[6] Sawhney, N. and Schmandt C., Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments, ACM Transactions on Computer-Human Interaction 7(3) (2000).
[7] Barry, A. and Elizabeth, M. The Future of Speech and Audio in the Interface, SIGCHI 26(4) (1994).
[8] Arons, B., SpeechSkimmer: A System for Interactively Skimming Recorded Speech, ACM Transactions on Computer-Human Interaction 4(1) (1997).
[9] Brewstwer S. A., Using Nonspeech Sounds to Provide Navigation Cues, ACM Transactions on Computer-Human Interaction 5(3) (1998).
[10] Hurst A., et al. The Sense Lounger: Establishing a Unicomp Beachhead in Elders' Homes, in: Proceedings of CHI, 2005.
*Jodi Forlizzi, Private communication.