THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.21, No.4E 2002. 12. pp. 170~177

Dialog System Using Multimedia Techniques for the Elderly with Dementia

Sung-Ill Kim*, Hyun-Yeol Chung**
*Center of Speech Technology, State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science & Technology, Tsinghua University, China
**School of Electrical Engineering and Computer Science, Yeungnam University, Korea
(Received 2 September 2002; accepted 10 December 2002)

Abstract

The goal of the present research is to improve a quality of life of the elderly with a dementia. In this paper, it is realized by developing the dialog system that is controlled by three kinds of modules such as speech recognition engine, graphical agent, or database classified by a nursing schedule. The system was evaluated in an actual environment of a nursing facility by introducing it to an older male patient with dementia. The comparison study between dialog system and professional caregivers was then carried out at nursing home for 5 days in each case. The evaluation results showed that the dialog system was more responsive in catering to needs of dementia patient than professional caregivers. Moreover, the proposed system led the patient to talk more than caregivers did.

Keywords: *Dialog system, Speech recognition, Dementia, Caregiver, Virtual conversation*

I. Introduction

As it has become an aging society both in developed and in developing countries, the overall quality of life (QOL) of the elderly is getting more and more important issue in reality. Particularly, recovering a good health physically and mentally from diseases is an essential purpose of the life for all elderly. Therefore, such related technologies and secure environments have been designed for independent living and social participation of older persons suffering from diseases.

The present study aims to improve the mental health-related QOL of the elderly with a dementia[1-4]. For the method of realizing it, the mutual interaction through conversation was used for the mental or emotional stability of the older people with dementia. Moreover, the caregiving stress at home often leads to problems in caregivers' mental and physical health because of the behavioral problems of their family member with dementia. Therefore, this study also aims to improve the QOL of family caregivers by lightening their nursing loads to some degree in long-term care[5,6].

On the basis of these social backgrounds, we have developed the dialog system[7-11] based on the techniques of multimedia application. The system is focused on a natural interaction between dementia patient and system through spontaneous speech, without using any other interfaces, such as a keyboard or mouse, etc. Therefore, the system was designed to be a good conversational partner to dementia patients whenever they need. As a consequence, the emotional stability might be recovered by mutual communication, thus having effects of rehabilitation. Furthermore, the system might be helpful for the

Corresponding author: Hyun-Yeol Chung (hychung@yu.ac.kr)
School of Electrical Engineering and Computer Science, Yeungnam University, 214-1, Gyungsan, Gyungbuk, 712-749, Korea

nursing works of family caregivers at home or professional caregivers at nursing facilities, particularly when they have a great pressure of work.

The proposed system should be optimized to only one dementia patient because of the diverse symptoms of dementia. In this study, therefore, subject has been confined to one aged person with a severe dementia. Moreover, the emotional state of the subject in the course of conversation was examined by analyzing the quantity and the contents of conversation.

II. Preliminary Investigation

2.1. Observation of Behavioral Patterns of Subject

The behavioral patterns of daily life of dementia patient were examined at nursing facility for a relevant design of dialog system. For this study, the official approvals from the ethics committee of nursing facility as well as from family members of subject were first obtained in advance. The subject was 72 years old male patient with a vascular dementia. He has received the day-care service at nursing home from 8:30 a.m. to 3:30 p.m. on weekdays. Particularly, the preliminary investigation focuses on his behaviors at nursing facility after lunch when he becomes mentally unstable.

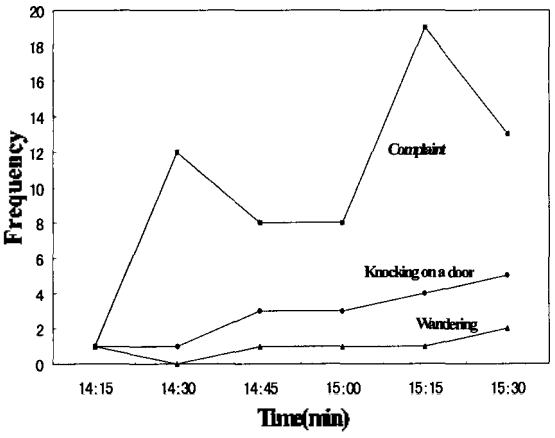

Figure 1. Average frequencies of three different behavioral patterns of subject, which were examined for 90 minutes after lunch during 3 days at nursing facility.

The behaviors and utterances of subject during the day-care service were taken by a video camera. Figure 1 shows the typical observation results between 2:00 p.m. to 3:30 p.m. when subject has the most uncomfortable emotional state. It shows the average frequencies of behavioral patterns such as complaint, knocking on a door, or wandering, all of which were investigated for 90 minutes after lunch during 3 days.

It was noticed that subject had the most frequent behavioral pattern in complaint among three different ones, increasingly demanding what someone caters to his needs. Since dementia patient like subject is very sensitive to changes in resident environment or life schedule and dementia usually has a disorder of memory, these factors can cause an extreme disappointment or frustration, occasionally developing an agitation or mental depression. Accordingly, it should be taken into account that the proposed system is designed in order to become a conversational partner as a good listener, sometimes leading him to talk, so that it might help him to restore a comfortable emotional state.

2.2. Criteria for Selecting Dialog Database

The frequently used words of subject, which were based on the results of the examination mentioned above, were surveyed for building dialog database of system. In general, the elderly with dementia often finds it hard to remember the meaning of words used in their daily lives, or to think of words they want to say. In this study, therefore, we selected the most relevant dialogs as to how to talk to dementia patient for a virtual communication between patient and system, according to the following criteria on the basis of the experiences of professional caregivers or occupational therapists.

First, dialog system should try to use words that do not make agitation or frustration worse for the elderly with dementia. Second, system should be designed to call by name or intimate nickname. Third, system is necessary to be designed to speak slowly and distinctly, and use familiar words and short sentences mixed with regional dialects that are familiar to the patients. Fourth, system should be designed to keep things positive, for instance,

offering positive choices like "Let's take a rest by the time your family comes here", or "Lunch would be served with your favorite things". Finally, system should be designed to ask simple questions that can be answered with yes or no or one-word answers, when it is difficult to know what they want.

Under these criteria, the several sorts of dialog database collected by two female professional caregivers whose voices were particularly more responsive for subject than the voices of others. When building the database, more importantly, the changes of demands in time should be taken into account to make him retain emotional peace. Moreover, the system needs to have a supplementary function of announcing the necessary information on the schedule of day-care service.

III. Outline of Dialog System

The dialog system for dementia patient is required to be equipped with the following major functions based on the outcomes of the above-mentioned preliminary examination.

(1) The techniques of the command speech recognition with speaker adaptation were used for speech recognition module. In this study, 50 different words and phrases were used for recognition. It is due toa dialog pattern of the user with dementia of usually speaking with the limited and repeated utterances, which have been found in the preliminary observation of the utterance patterns of dementia patient.

(2) For a response of system, it is organized to reply relevantly by speech synthesis or recorded voices to user's demands. In this case, a graphical interface with a virtual caregiver is synchronized with the responses.

(3) System provides functions of chiming with user or making agreeable responses to demands. This function enables system to interact with user more smoothly and naturally.

(4) System captures the time when the incoming voice signals are given to speech recognition module. At next step, it searches the most suitable response out of database at the moment, which has been classified into the time schedule of nursing facility.

(5) In case user asks back, system is designed to make the same response as the previous one, because dementia user has a tendency of reconfirming the answers at a frequent rate. In the preliminary investigation, it has been revealed to be very useful for natural interaction.

(6) When system is responding, the incoming signals are restrained to prevent it from becoming oversensitive.

(7) System has adaptability to switch the situation of day-care service to different environments such as short-stay service at nursing facility or home-care service. It is easily realized by substituting for the desired dialog database suitable for the new situation.

(8) System is adaptable to different users with the different symptoms of dementia by registering their individual utterance patterns.

Figure 2 shows the main frame of dialog system with the functions mentioned above. Basically, the proposed system chiefly depends on the function of speech recognition for interaction. However, its high performance is not always guaranteed owing to the ambient noisy environments and slurred speech of the elderly with dementia. In this study, therefore, the additional functions such as usage of database classified by a timetable of

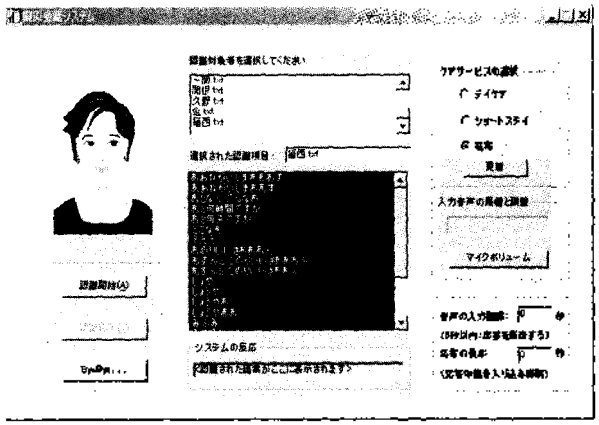

Figure 2. Main frame of dialog system with functions of candidate list for speech recognition, selective option of different services and users, and graphical user interface.

nursing home or chiming with user supplement a shortcoming of speech recognition for maintaining natural interaction. Therefore, this enables system to promote natural conversation between system and user even when system fails in recognizing speech of user correctly, or when the incoming speech is not registered in candidate lists. As a result, user might feel as if dialog system hears and understands his words attentively, just like familiar caregivers at nursing facility or family members.

Figure 3 illustrates a schematic diagram to form interaction between dialog system and dementia user. The interaction manager controls the three kinds of major modules such as speech recognition engine[3,4], graphical agent, or database classified by nursing service schedule. When user utters, the input speech signals are first captured by speech recognition engine module. The recognized results are then given to the interaction manager, which captures the present time of acquiring input signals, to select the most likely and suitable reply out of database at the time interval. It then makes a response with recorded voices, simultaneously synchronizing them with a lip of virtual caregiver image. In this case, the function of detecting input signals is suspended during the response of system for the prevention of oversensitive response to input signals, and then ready for detecting the next input speech. Moreover, system is designed to make the same response again if the input voices are captured within an arbitrary preset time.

IV. Evaluation

4.1. Operating Experiments

In the preliminary investigation, it was noticed that subject demanded something more frequently after lunch. Therefore, the dialog system was set after lunch when he was in the most uncomfortable emotional state. In experiments, we used recorded voices of caregivers for natural interaction, which had been proved to be more responsive for him than using synthesized ones. The operating experiments of system were performed in the main hall of nursing facility, which was a relatively noisy because it was the place where nursing home residents have rests or meals. Figure 4 shows a virtual conversation between dementia patient and dialog system using a wireless microphone (WT-1110, TOA, Japan) for free behaviors of subject.

The comparative study was examined where one of the evaluations was performed with dialog system for 90 minutes after lunch during 5 days. In this case, the interruption of caregivers was restricted during the experiment. The other evaluation was performed during the other 5 days without system, where caregivers were allowed to do nursing activities freely whenever subject demanded. Each evaluation test was conducted every second day. During the period of the experiments, occupational therapists observed subject's reactions to

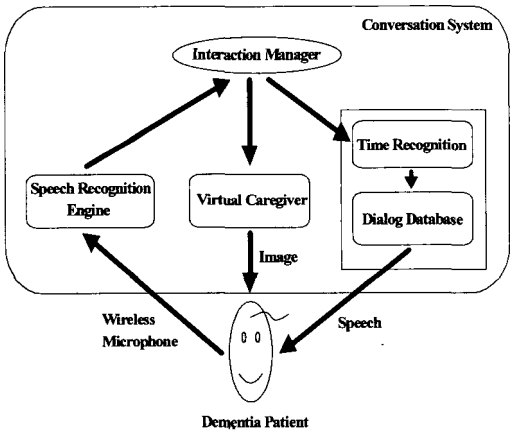

Figure 3. Schematic diagram of interaction between dialog system and dementia user.

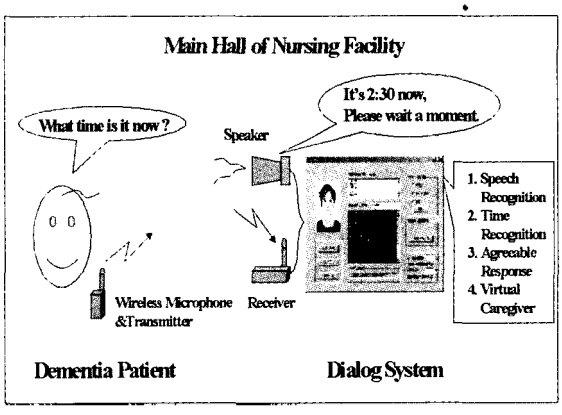

Figure 4. Virtual conversation between dementia patient and dialog system in the main hall of nursing facility.

system and caregivers, simultaneously photographing his behaviors using a video camera.

4.2. Results

For comparative analysis, we examined the frequency of the demands of subject, and the frequency of each handling response to the demands for both system and caregiver as well as the corresponding reactions of subject. Table 1 shows the frequencies of three kinds of components with and without dialog system, respectively. In this table (a) indicates the frequencies of demands (n=389), response of caregivers (n=82), and reaction of subject (n=26), which were investigated for 90 minutes of each day during 5 days when system was not given. On the other hand, (b) shows the frequencies of demands (n=593), response of system (n=468), and reaction of subject (n=233), which were also investigated under the same condition when system was given.

Table 1. Frequency comparison. (a) The frequencies of the demands of subject, the response of caregivers, and the reaction of dementia subject when system was not given. (b) The frequencies of the demands of subject, the response of system, and the reaction of dementia subject when system was given.

	Frequency of Demands of Subject	Frequency of Response of Caregivers	Frequency of Reaction of Subject
1'st day	111	9	0
3'rd day	122	27	2
5'th day	17	0	0
7'th day	77	29	13
9'th day	62	17	11
Total	389	82	26

(a) Without system

	Frequency of Demands of Subject	Frequency of Response of System	Frequency of Reaction of Subject
2'nd day	65	41	13
4'th day	136	94	57
6'th day	134	106	52
8'th day	143	129	66
10'th day	115	98	45
Total	593	468	233

(b) With system

For a quantitative evaluation, we induced the response degree as a ratio of each response frequency for both caregivers and dialog system to the demand frequency of subject. Figure 5 shows the comparison of response degrees for both caregivers and dialog system to the demands of subject. It was found that caregivers showed the responses by 21% to the demands of subject without system, whereas dialog system showed 79%. Judging from the comparison results, although caregivers are possible to handle subject's demends by 21%, they were not still responsive by the other 79%. On the other hand, the handling of system reached 79% in which it had only the other 21% of non-response owing to the insensitivity of detecting incoming speech signals. In comparison of the both graphs, consequently, it revealed that system was more responsive by 58% than caregivers.

Figure 6 shows the comparison of the reaction degrees of subject to each response of both caregivers and dialog

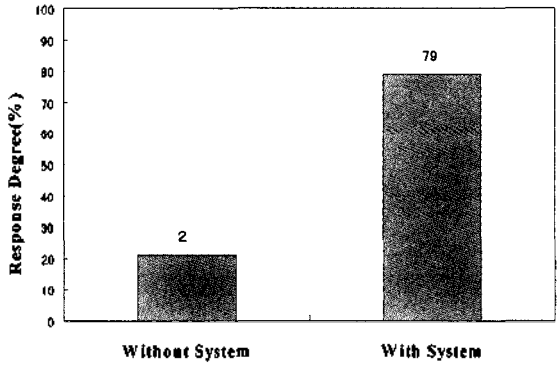

Figure 5. Comparison of the response degrees of both caregivers (without dialog system) and dialog system to the demands of subject.

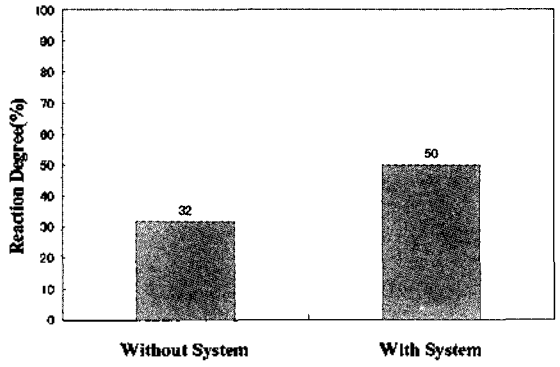

Figure 6. Comparison of the reaction degrees of subject to each handling of caregivers and dialog system.

Table 2. Frequencies of the reactions of subject such as affirmation, asking back once or twice when system was given and not, respectively.

Date	Asking Back (once)	Affirmation	Asking Back (twice)
1'st day	0	0	0
3'rd day	2	0	0
5'th day	0	0	0
7'th day	12	1	0
9'th day	6	5	0
Total	**20**	**6**	**0**

(a) Without system

Date	Asking Back (once)	Affirmation	Asking Back (twice)
2'nd day	8	4	1
4'rd day	29	16	12
6'th day	16	24	14
8'th day	28	25	13
10'th day	14	22	9
Total	**95**	**91**	**49**

(b) With system

system. The reaction degrees were induced by a ratio of each response frequency for both caregivers and dialog system to the reaction frequency of subject. When system was not given, subject showed reactions by 32% to the handling of caregivers. When system was given, on the other hand, he showed 50% to the handling of system. Thus, It was shown that subject was more interactive with system by 18% than with caregivers.

Although the above results show that subject had more frequent interaction with system, it never indicates that more natural conversation is always built. To evaluate how effectively system can maintain a smooth interaction with dementia user, therefore, it is important to study what kind of emotional states (for example, emotional burden or peace) user has in a process of interaction with dialog system. The professional therapists first examined subject's reactions to each handling of both caregivers and system by analyzing the contents of videotapes, as a further analysis of figure 6. The reactions of subject were then divided into several components. As a first component, subject showed reactions of asking back to get or

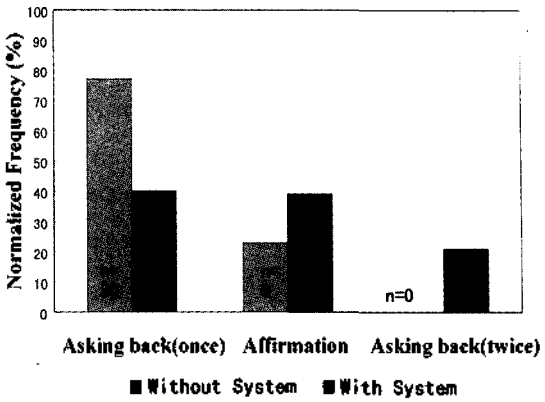

Figure 7. Comparison of each normalized frequency of three different reactions (n=frequency).

reconfirm some agreements, particularly when he heard necessary information from caregivers or dialog system as information provider. As a second component, he showed reactions of asking back twice, particularly when the answers to his question were dissatisfactory or different from his expectation. Finally, he showed affirmative reactions with simple answers such as 'yes' or 'I see'.

By means of this analysis, we could obtain the frequencies of three different reactions such as affirmation, asking back once or twice, which were indicated in table 2. Figure 7 shows the comparison of each normalized frequency of the reactions. The notable differences were not shown in the components of asking back once and affirmation. On the other hand, subject had no any reactions in the component of asking back twice to the handling of caregivers. However, he showed reactions of asking back twice to the handling of system. The reactions were shown particularly when the virtual conversation was not built smoothly or naturally, causing uneasiness or emotional instability.

V. Discussion

In the evaluation results, we could find that the speech recognition accuracies were degraded owing to subject's slurred voices mixed with regional dialects as well as his characteristic accents. Nevertheless, the supplementary functions such as chiming with user by making agreeable responses made conversation to be smooth, so that he

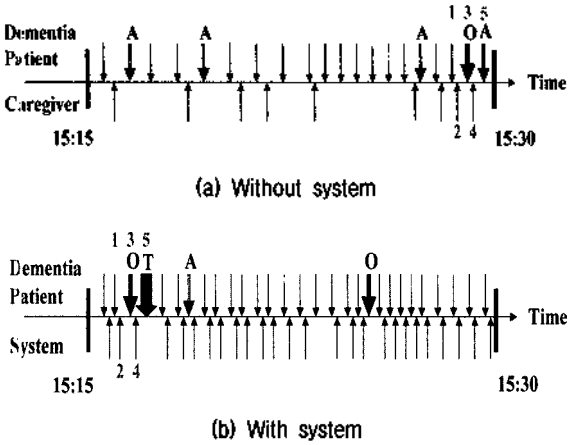

Figure 8. Examples of interaction between dementia patient and caregivers(a)/system(b).
(A: Affirmation, O: Asking back once, T: Asking back twice)

regarded system as a good listener. As shwn in comparison results in figure 5, system was more responsive in dealing with subject's demands or complaints than caregivers of busy nursing schedules. In addition, he had more active interaction with system than with caregivers as shown in figure 6. Therefore, system might be expected to lighten the loads of the nursing works of caregivers at nursing facilities or at home, by introducing system during their busy time, and thus the QOL of them might be improved as a subsidiary effect. Moreover, the rehabilitative effects might be achieved through a prompt response whenever the dementia patients need a conversation.

In this study, system was optimized to only one dementia patient because of the diverse symptoms of dementia. However, the proposed system has been already equipped with extensional functions to be adaptable to different dementia users by registering their individual utterance patterns. Besides, system has been designed to be flexible to different environments at nursing facility or at home.

From the evaluation results, it is found that we still have essential issues to solve. As illustrated in figure 7 and its typical examples in figure 8, subject had no any reactions of asking back twice to caregivers because interaction between them was natural and satisfactory. It implies that natural and timely conversation never gives some kind of emotional stresses or such burdens to dementia patients. When dialog system was introduced, on the other hand, it was noticed that there were occasional reactions of asking back twice. For example, he asks back twice (number 5 in figure 8(b)) because the answer of system (number 4) is unsatisfactory or unwanted to his asking back (number 3). It was mainly due to the failure in recognizing subject's speech correctly, thus causing unnatural conversation between them. It eventually developed an agitation or uneasiness so that he became stressful or depressive in interaction with system.

Moreover, there is one thing to remember from the experimental results. It is about roles of nonverbal communication such as emotional sympathy, body touch, etc., especially when talking with patients suffering from dementia. Actually, caregivers often talk with their patients while taking hands. Those nonverbal interactions also build a new type of mutual communication as well as human speech.

VI. Conclusion

This study aims to improve the QOL of the elderly who have been suffered from dementia. In order to realize this purpose, we have developed the system to respond with a natural dialog to their needs. As a result of the survey, we were able to draw two different conclusions. The one thing was that the proposed system was more responsive in catering to needs of dementia user than professional caregivers. The other thing was that dialog system caused more utterances of user than caregivers did. Moreover, the proposed system has shown a possibility to lighten nursing loads or burdens of caregivers at nursing facilities or at home.

References

1. Tomoeda CK, Bayles KA, Trosset MW, Azuma T, Mcgeagh A., "Cross-Sectional Analysis of Alzheimer Disease Effects on Oral Discourse in a Picture Description Task," *Alzheimer Disease & Associated Disorders*, 10 (4), 204-215, 1996.

2. Post, Stephen G., "The Moral Challenge of Alzheimer Disease," John Hopkins University Press, 1995.
3. Minthon L, Edvinsson L, Gustafson L., "Correlation Between Clinical Characteristics and Cerebrospinal Fluid Neuropeptide Y Levels in Dementia of the Alzheimer Type and Frontotemporal Dementia," *Alzheimer Disease & Associated Disorders*. **10** (4), 197-203, 1996.
4. Fox, Patrick, "From Senility to Alzheimer's Disease: The rise of the alzheimer's disease movement," *The Millbank Quarterly*. **67** (1) 58-102, 1989.
5. Kitwood, T., "Experience of Dementia, Aging and Mental Health," 1:13-22, 1997.
6. K. John, "The experience of dementia: a review of the literature and implications for nursing practice," *Journal of Clinical Nursing* 5: 275-288, 1996.
7. S. Nakagawa, A. Kai, T. Itoh and S. Kogure, "Speech recognition and understanding of spoken dialogue," *Proc. Int. Symposium on Spoken Dialogue*, 5-1-5-6, 2000.
8. H. Fujisaki, K. Shirai, S. Doshita, and S. Nakagawa, etl., "Overview of an intelligent system for information retrieval based on human-machine dialogue through spoken language," *Proc. Int. Conf. Spoken Language Processing*, I-70-73, 2000.
9. Rainer Stiefelhagen, Jie Yang, Alex Waibel, "Tracking Focus of Attention for Human-Robot Communication," *IEEE-RAS International Conference on Humanoid Robots -Humanoids 2001*, 22-24, 2001.
10. A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, etl., "Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech," *Computational Linguistics* 26 (3), 339-373, 2000.
11. H. Asoh, T. Matsui, J. Fry, F. Asano, and S. Hayamizu, "A spoken dialog system for a mobile office robot," *Proc. of Eurospeech'99*, 1139-1142, Budapest, September, 1999.

[Profile]

• **Sung-Ill Kim**

Sung-Ill Kim was born in Kyungbuk, Korea, 1968. He received his B.S. and M.S. degrees in the Department of Electronics Engineering from Yeungnam University, in 1997, and Ph.D. degree in the Department of Computer Science & Systems Engineering from Miyazaki University, Japan, in 2000. During 2000 to 2001, he was a postdoctoral researcher in the National Institute for Longevity Sciences, Japan. Currently, he has been working in the Center of Speech Technology, Tsinghua University, China. His research interests include speech/emotion recognition, neural network, and multimedia signal processing.

• **Hyun-Yeol Chung**

Hyun-Yeol Chung was born in Kyungnam, Korea, 1951. He received his B.S. and M.S. degrees in the Department of Electronics Engineering from Yeungnam University, in 1975 and 1981, respectively, and the Ph.D. degree in the Information Sciences from Tohoku University, Japan, in 1989. He was a professor from 1989 to 1997 at the School of Electrical and Electronic Engineering, Yeungnam University. Since 1998 he is a professor in the Department of Information and Communication Engineering, Yeungnam University. During 1992 to 1993, he was a visiting scientist in the Department of Computer Science, Carnegie Mellon University, Pittsburgh, USA. He was a visiting scientist in the Department of Information and Computer Sciences, Toyohashi University, Japan, in 1994. He was a principle engineer, Qualcomm Inc., USA, in 2000. His research interests include speech analysis, speech / speaker recognition, multimedia and digital signal processing application.