Interactive Rehabilitation Support System for Dementia Patients

Sung-ill Kim*

Abstract

This paper presents the preliminary study of an interactive rehabilitation support system for both dementia patients and their caregivers, the goal of which is to improve the quality of life(QOL) of the patients suffering from dementia through virtual interaction. To achieve the virtual interaction, three kinds of recognition modules for speech, facial image and pen-mouse gesture are studied. The results of both practical tests and questionnaire surveys show that the proposed system had to be further improved, especially in both speech recognition and user interface for real-world applications. The surveys also revealed that the pen-mouse gesture recognition, as one of possible interactive aids, show us a probability to support weakness of speech recognition.

Keywords : Interactive System, Rehabilitation, Dementia, Speech Recognition, Face Verification.

1. Introduction

The number of people with dementia is expected to rise with the aging population steadily increasing. Dementia causes a progressive decline in intellectual functioning, communicative abilities. Hence, it makes it more difficult to communicate with dementia patients because of such a loss of memory and cognitive function. In consequence, this often leads to a mental stress or a burden to professional therapists or family caregivers, in the nursing facilities or at home, as outcomes of long-term nursing activities.

The ultimate goal of our approach is to raise the quality of life(QOL) and the rehabilitative will of the elderly with dementia. This study also aims to improve the QOL of their caregivers such as professional therapists or family members, by lightening their nursing loads to some degree in long-term care. For realizing it, the virtual interaction through conversation has been used for mental or emotional stability of the older people with dementia. In addition, the system based on the interactive method might be helpful for caregivers to do their healthcare services.

On the basis of these social backgrounds, several interactive systems[1,2,3,4,5] in terms of the rehabilitation of dementia patients have been developed so far. One of the systems is based on using a touch screen interface[1] as an interaction. Some of another systems is based on a multimedia computer system[2]to support communication, and on a computer-basedmemory training[3]. Bothour previous[4,5] and present systems have been focused on the virtual interaction between user and system. As the interactive techniques, the previous study had been developed by simply using speech recognition, whereas the present study included face and pen-mouse gesture recognition as well as speech recognition.

As a prior condition of the interactive system, especially, we focused on some basics, that is to say, easy to use and intuitive to access it. Figure 1 shows an overview of the proposed interactive system for dementia patients as our preliminary study, including two different modules that supplement virtual interaction. The first module is the face recognition which is used for identifying specific users in order to provide them with the most optimal services. The second module is the pen-mouse gesture recognition, especially for easy use of caregivers, which is used for making up for the faults of speech recognition which may fail to give satisfactory results due to a recognition mismatch.

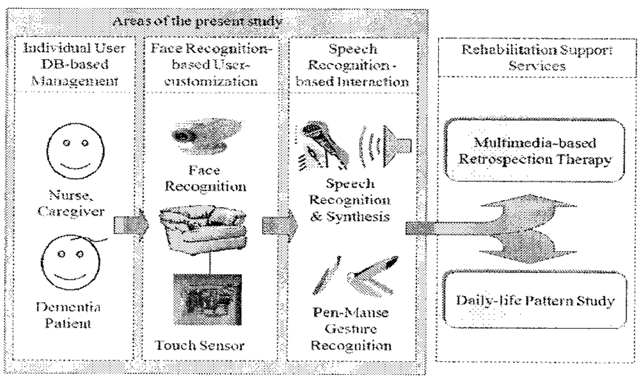

Fig.1. Concept of the interactive system for dementia patients, based on the recognition modules such as speech recognition, pen-mouse gesture recognition, and face recognition.

---

* Dept. of Electronic Eng, Kyungnam Univ, Korea
투고 일자 : 2010. 5. 31 1차 수정 : 2010. 7. 23
2차 수정 : 2010 7. 28 게재확정일자 : 2010. 7. 29
※ This work was supported by the Korea esearch Foundation Grant funded by the Korean Government(KRF-2008-331-D00431).

These two modules enable the proposed system to be easily used for caregivers as well as patients. Therefore, the system was designed to become a good interactive partner for dementia patients whenever they need. Through this, the emotional stability of the patients might be recovered by the virtual interaction, thus having effects of rehabilitation. Furthermore, the system might be helpful for the nursing works of family caregivers at home or professional caregivers at nursing facilities. In order to do this, the proposed system should be optimized to dementia patients with diverse symptoms of dementia. In this study, therefore, the system processes an interaction based on individual user's database.

2. The Interface for Interaction

In speech recognition systems, the HMM(Hidden Markov Model) has been well-known and widely used based on the statistical method of characterizing the spectral properties of the frames of a pattern. In this study, we used HM-Net(Hidden Markov Network)[6] which is an efficient representation of context-dependent phonemes for speech recognition. The HM-Net, which has various state lengths and shares their states one another, is automatically generated by SSS(Successive State Splitting)[7,8]. The SSS is an iterative algorithm that is able to simultaneously determine and estimate an optimal set of allophones, an optimal state sharing architecture, and optimal parameters with the maximum likelihood criterion. According to which has the largest divergence between its two mixtures, the state is split on the contextual and temporal domains, and the one giving greater likelihood is then chosen. The affected states are retrained using the Baum-Welch algorithm[9,10]. The above procedure is iteratively performed until getting to pre-defined number of states.

The PDT-SSS(Phonetic Decision Tree-Successive State Splitting)[11] based on the SSS algorithm is a powerful technique to design topologies of tied-state models, and is possible to generate highly accurate HM-Net. The final result of state splitting is a network of states that efficiently represents a collection of context-dependent models. In contrast to the training process of the existing HMM, the architecture of the models can be automatically optimized according to the duration of utterances. As a result, the number of states in vowels increases more than that of states in consonants in terms of the architecture.

Figure 2 shows an overall schematic of HM-Net speech recognition system. In case input speech signals are given to the system, the acoustic features are first extracted for pre-processing, and then given to the first and the second pass search modules that use tree-structured lexicon, HM-Net Triphones, and semantic grammars. The HM-Net based speech recognizer had been proved that it showed better performance than the conventional HMM based recognition system in the experiments of phoneme, word, and continuous speech recognition[12,13].

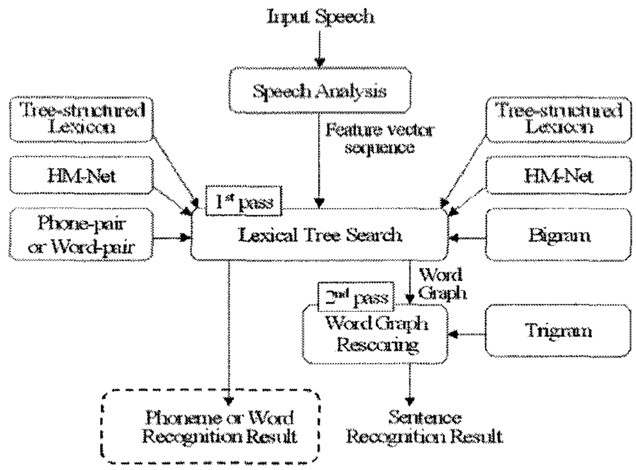

Fig.2. Overall schematic of HM-Net speech recognition system.

The interaction of the proposed system is conducted by the module of the speech recognition. However, the performance of speech recognition might be degraded in noisy environments, especially when playing multimedia players at the same time. In this study, therefore, the complementary function of speech recognition was adopted by incorporating the module of pen-mouse gesture recognition into the system. Fig. 3 shows simple examples for the practical applications based on the gesture recognition which was designed and implemented using Visual C++. A wireless pen-mouse(i-pen bluetooth, Finger System Inc.) was used for this application. The gesture recognition was based on an simple algorithm discriminating direction angles. Six kinds of mouse gestures were designed for controlling multimedia player, whereas one of them is used for the case of mismatch of speech recognition to turn back to a previous command. Therefore, this module can be used as a complementary device of speech recognition.

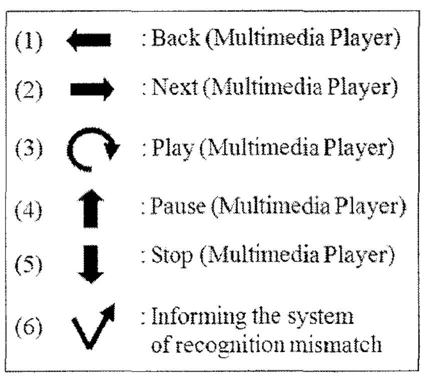

Fig. 3. Application examples of gesture recognition using a wireless pen-mouse.

3. The Interface for Face Verification

In the present study, the interface for face verification is based on the robust digital image processing algorithms which include the modules of advanced face localization, enrollment and matching. In the process of verification, it first processes the input face image, extracts features and then writes them to the database. In the module of face enrollment with a feature generalization, particularly, it generates the collection of the generalized face features from a number of the face templates of the same person. Each face image is first processed and the features are then extracted. In the next step, the collections of features are analyzed and combined into one generalized feature collection which is then written to the database for a matching process. The interface provides a high level of reliability if the face enrollment mentioned above is successfully carried out. In the module of face matching, on the other hand, it performs the matching process between the input face image and the face templates stored in the database.

This interface was incorporated into the proposed system for a user customization. In this study, we used the VeriLook SDK[14] for real applications. Figure 4 shows the user interface of the face verification realized by VC++(ver.6.0). The main application window of the interface has a layout with four panes where two top panes are used for image display and two bottom panes are used for message logging. The face detection pane is used to display the still images, the videos, or the result of face detection overlaid on image. The matching/enrollment pane is used to display images enrolled in face database or used for matching. The application log pane is used for both the system information and the progress messages. The matching results pane is used for listing ID of the subject enrolled in database, most similar to the matched image. Subjects are considered similar if their similarity value exceeds matching threshold set.

Fig.4. Interface for face verification (No. 1: face detection pane, No. 2: application log pane, No. 3: matching/enrollment pane, No. 4: matching results pane)

4. The Proposed Interactive System for Dementia Patients

The proposed system was built by integrating the interfaces for both interaction and face verification, mentioned in the previous chapters. Figure 5 shows the flow diagram of the entire processing based on the proposed system which is operated in real time. It shows how to build the interaction between user and system.

If user sits on sofa, the system catches signals from the touch sensor laid on the sofa and then automatically activates the face identification engine to detect a face area. If the system recognizes who is sitting on the sofa, it adapts itself to the specific user by searching and adopting the individual database to support the most optimal services for the user. It then activates the speech recognition engine where the virtual interaction between user and system is built using speech recognition and synthesis(i-Talk SDK version 2.2). In case speech recognition is activated, it is ready to provide the specific user-oriented services. It can provide the user, who is dementia patient, with the rehabilitative support services such as multimedia based retrospection therapies or daily-life pattern studies. it can be done by retrieving multimedia database(favorite music, private photos or videos) using speech recognition or pen-mouse gesture recognition, with the help of caregivers. In the proposed system, the list of the registered recognition candidates can be automatically updated according to the corresponding recognition results.

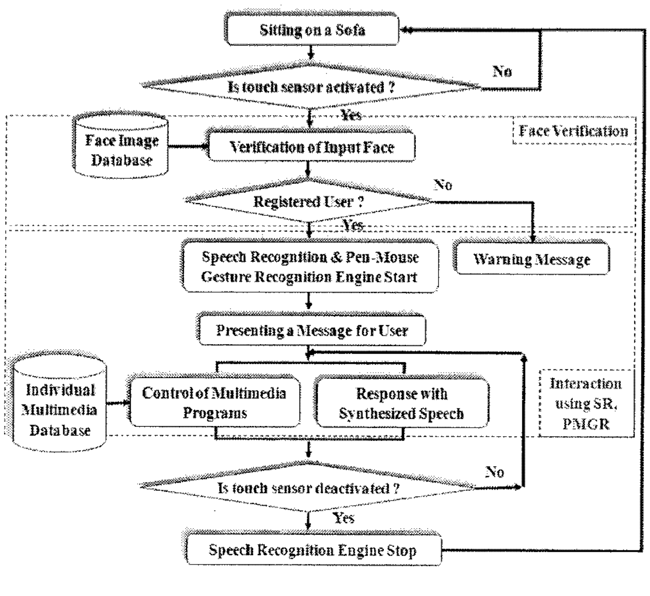

Fig.5. The flow diagram of building interaction between user and system

Figure 6 shows the main user interface for the proposed interactive system made by VC++(ver.6.0), and by integrating three different interfaces such as speech recognition, pen-mouse gesture recognition and face verification. The proposed system has another functions for supporting an interaction. It is possible for user to control multimedia-based application programs such as video, MP3 and CD players. In addition, the touch sensor shown in figure 7 was connected to the system through a print-port control. It was attached on the sofa so that the input signals obtained from the sensor can activate or deactivate the main system in case someone sits on the sofa.

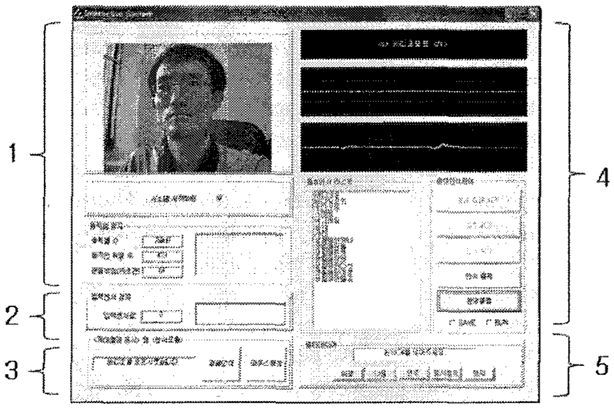

Fig.6. Main user interface for the proposed interactive system. (No. 1: the interface of video processing, No. 2: the interface of touch sensor detection, No. 3: the interface of face recognition and pen-mouse gesture recognition, No. 4: the interface of speech recognition, No. 5: the interface of multimedia control)

Fig.7. Touch sensor(left) attached on the sofa(right)

5. Experimental Results

The speech signals for both HMM training and recognition process were sampled at 16kHz, quantized at 16 bits, pre-emphasized with a transfer function of $(1-0.97z^{-1})$, and processed to extract acoustic features using a 25ms Hamming window with a 10ms shift. The feature parameters consisted of 39th order LPC Mel Cepstrum coefficients, including a normalized log-power, the first and the second order delta coefficients. Table 1 shows the analysis of speech signals.

Table 1 shows the preprocessing of speech signals to extract feature parameters. In this study, the database of ETRI(Electronics and Telecommunications Research Institute, Korea) was used in the training process for speech recognition.

Table 1. Preprocessing of speech signals for speech recognition.

Item	Contents
Sampling rate	16kHz , 16bits
Pre-emphasis	0.97
Window	25 ms Hamming window
Frame period	10 ms
Feature Parameter	13th order LPC MEL Cepstrum +13th order ΔLPC MEL Cepstrum + 13th order ΔΔLPC MEL Cepstrum = Total 39th order LPC MEL Cepstrum

As preliminary experiments of the interactive system for dementia patients, total 25 male college students were participated in the evaluation of the proposed system. For examining the human performance on the accuracies of the system, we first showed them a demonstration of how to use and operate the system, and made them to use it for themselves. Table 2 shows the average recognition accuracies in each module such as speech recognition, pen-mouse gesture recognition and face recognition. For the evaluation of speech recognition incorporated into the proposed system, total 475 utterances(25 users * 19 utterances) were used. The evaluation was performed in the laboratory environments with the noises such as computer cooling fan or buzz of voices. In experiments, we adopted speech recognizer with 2,000 states and 4 mixtures per state. For the evaluation of pen-mouse gesture recognition, total 150 pen-mouse gestures (6 gestures * 25 users) were used. For the evaluation of face recognition, on the other hand, the facial images of 25 male college students were first registered in database and the recognition test in each user was then conducted.

Table 2. Accuracies of three recognition modules such as speech, pen-mouse gesture and face recognition.

Module	Accuracy(%)
Speech Recognition	(330/475)*100=69.5
Pen-mouse Gesture Recognition	(139/150)*100=92.7
Face Recognition	(24/25)*100=96.0

As the evaluation using questionnaire, all participants marked ranks from 1-point(very difficult, never useful) to 5-point(very easy, very useful) about both how easy and how useful they thought the system was to use. As shown in the questionnaire survey of figure 8, the results which took note of the frequencies of the point 1 and 2(difficult to use as well as not useful) show that the proposed system needs further improvements, especially in speech recognition and user interface for real world applications. It is mainly due to the degradation of speech recognition influenced by ambient noises, diverse speaking rates, and speaking styles of users. However, the surveys also revealed that the recognition module for pen-mouse gestures showed us a realizable possibility that made up for the weakness of speech recognition.

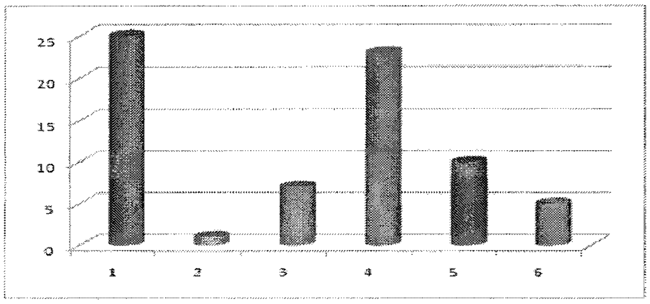

Fig.8. Modules or Interfaces that should be improved (No. 1: speech recognition, No. 2: face recognition, No. 3: real-time processing, No. 4: user interface, No. 5: pen-mouse gesture recognition, No. 6: etc.)

6. Conclusion

A preliminary study on the interactive method for dementia patients was investigated. The proposed system made use of the integrated method based on speech, face and pen-mouse gesture recognition. As a result of questionnaire survey, it was revealed that the proposed system needs to be improved, especially in speech recognition and user interface. The accuracy of speech recognition is unsatisfactory due to ambient noises, diverse speaking rates, and speaking styles of users. Nevertheless, it seems that the alternative device, such as recognition module for pen-mouse gestures, can improve an imperfection of speech recognition.

As future works, the present system should be conducted on more natural methods for interaction so that the future system would allow users to feel more natural in virtual interaction. In addition, the system should be applied to both dementia patients and their caregivers in the real field of nursing facilities.

References

[1] P. Riley, N. Alm, A. Newell, "An interactive tool to promote musical creativity in people with dementia", Computers in human behavior, Vol.25, No.3, pp.599-608, 2009.

[2] A. Astell, N. Alm, G. Gowans, "Involving older people with dementia and their carers in designing computer based support systems: some methodological considerations", Universal access in the information society, Vol.8, No.1, pp.49-58, 2009.

[3] M.Schreiber, A. Schweizer, K. Lutz, K. T. Kalveram, L. Jäncke, "Potential of an Interactive Computer-based Training in the Rehabilitation of Dementia: An Initial Study", Neuropsychological rehabilitation, Vol.9, No.2, pp.155-167, 1999.

[4] S-I. Kim, B-C. Kim, "Dialog System based on Speech Recognition for the Elderly with Dementia", The Journal of the Maritime Information and Communication Sciences, Vol.6, No.6, pp.923-930, 2002.

[5] S-I. Kim, H-Y. Chung, "Dialog System Using Multimedia Techniques for the Elderly with Dementia", The Journal of the Acoustical Society of Korea, Vol.21, No.4E, pp.170-177, 2002.

[6] M. Suzuki, S. Makino, A. Ito, H. Aso and H. Shimodaira, "A new HMnet construction algorithm requiring no contextual factors," IEICE Trans. Inf. & Syst., Vol.E78-D, No.6, pp.662-669, 1995

[7] M. Ostendoft and H. Singer, "HMM Topology design Using Maximum Likelihood Successive State Splitting," Computer Speech and Language Vol.11, pp.17-41, 1997.

[8] J. Takami and S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. of ICASSP'92, Vol.1, pp.573-576, 1992.

[9] X. Huang, A. Acero, and H.W. Hon, "Spoken language processing: a guide to theory, algorithm, and system development," Prentice Hall, 2001.

[10] L. Rabiner, and B.H. Juang, "Fundamentals of speech recognition," Prentice-Hall International, Inc. 1993.

[11] T. Hori, M. Katoh, A. Ito and M. Kohda, "A Study on HM-Nets using Decision Tree-based Successive State Splitting," Proc. of ICSP'97, Vol.1, pp.383-387, 1997.

[12] S-J. Oh, K-D. Kim, D-G. Roh, S-O. Wi, M-G. Song, H-Y. Chung, "A Study on Performance Evaluation of Hidden Markov Network Speech Recognition System(Korean)," KISPS, pp.30-39, 2003.

[13] S-J Oh, C-J Hwang, B-K Kim, H-Y Chung, "Performance Evaluation of HM-Nets Speech Recognition System using the Large Vocabulary Korean Speech Databases," Proc. of Kyushu-Youngnam Joint Conference on Acoustics, pp. 49-52, 2003.

[14] VeriLookSDK(Software Developer's Kit) version 2.0, Neurotechnologija, www.neurotechnologija.com/.

Sung-Ill Kim

Feb. 1994 : B.S. Dept. of Electronic Eng., Yeungnam Univ., Korea
Feb. 1997 : M.S. Dept. of Electronic Eng., Yeungnam Univ., Korea
Mar. 2000 : Ph.D. Dept. of Computer Science & Systems Eng., Miyazaki Univ., Japan
Apr. 2000—Mar. 2001 : Researcher, National Institute for Longevity Sciences, Japan.
Apr. 2001—Feb. 2003 : Researcher, Center of Speech Technology, Tsinghua Univ., China.
Mar. 2003—Feb. 2006: Full-time lecturer, Div. of Electrical & Electronic Eng., Kyungnam Univ., Korea
Mar. 2006—Feb. 2010 : Assistant professor, Dept. of Electronic Eng., Kyungnam Univ., Korea
Mar. 2010—Current : Associate professor, Dept. of Electronic Eng., Kyungnam Univ., Korea
[Research Interests]
Speech Recognition, Speech Processing, Image Processing, Multimedia Signal Processing