Design of Face-to-Face Multilingual Communication Environment for Illiterate People

Mai Miyabe[1] and Takashi Yoshino[2]

[1] Graduate School of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama, Japan
s085051@sys.wakayama-u.ac.jp
[2] Faculty of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama, Japan
yoshino@sys.wakayama-u.ac.jp
http://www.wakayama-u.ac.jp/ yoshino/

Abstract. In the medical field, a serious problem exists with regard to communication between hospital staffs and patients. Currently, although a medical translator accompanies a patient to medical care facilities, round-the-clock or emergency support is difficult to provide due to increasing requests. The medical field has high expectations from information technology. Therefore, we have developed a support system for multilingual medical reception termed M^3. We have installed our system in the Kyoto City Hospital in Japan. However, we found that our system cannot provide support to illiterate people. If an illiterate person and another person speak different languages, it is difficult the other person to communicate face to face with the illiterate person while explaining the meaning of texts shown on the display of the support system. This is one of the problems specific to the multilingual communication. There is a need to solve this problem. Therefore, we have developed a method to provide support to illiterate people engaging in multilingual face-to-face communication. We use a text-to-speech function implemented using a selector switch to provide support to illiterate people in performing operations using a touch screen. We performed an experiment to examine the effect of the proposed method. The results of the experiment are as follows. (1) From the results of the questionnaire, we find that the subjects are able to operate the selector switch easily. Therefore, we conclude that the method using the selector switch has little effect on the operation of the system. (2) Retrieval time using the text-to-speech function is five times that using the normal operation. We need to consider a structure that can retrieve the required information easily if many readings of texts are required.

Keywords: Parallel texts, Multilingual communication, Medical field.

1 Introduction

Opportunities for multilingual communication in Japan have increased due to the increase in the number of foreigners in Japan. When people communicate in their

N. Aykin (Ed.): Internationalization, Design, LNCS 5623, pp. 283 – 292, 2009.

nonnative language, the differences in languages prevent mutual understanding among communicating individuals. Differences in languages have to be overcome in order for multilingual communication to occur. To overcome the language barrier in communication, machine translation is used for communication using native language. We have conducted research on providing support for multilingual communication in the medical field. Currently, a medical translator accompanies a patient to medical care facilities, and the requests for medical translators to accompany patients are increasing. However, medical translators cannot provide support in cases in which round the clock support is required or in case of emergencies. In the medical field, a system that supports accurate multilingual communication is required. In the medical field, in particular, accurate translations are very important. Medical care directly impacts both human life and health. Despite recent advances in machine translation technology, it is still very difficult to obtain highly accurate translations. Inaccurate translation adversely affects communication, and an incorrect machine translation can cause serious problems.

We have developed a support system for multilingual medical reception termed M^3 [1]. Users operate M^3 using a touch screen and receive text-based support. M^3 provides reliable communication between a hospital staff member and a patient using accurate translations called parallel texts. However, in Japan, there are foreigners that are illiterate in Japanese or natives that are illiterate in their native language [2]. According to statistics, there are many countries worldwide that have a low rate of literacy [3]. In some countries, illiterate people account for more than 70% of the adult population. The actual illiterate population can be estimated to be larger because most countries estimate their illiterate population as being lower than it really is. If an illiterate person and his/her conversational partner speak the same language, the conversational partner can provide support to the illiterate person by a verbal explanation. On the other hand, if they speak different languages, it is difficult for the conversational partner to provide information on texts to the illiterate person by a verbal explanation. Therefore, illiterate people engaging in multilingual communication face problems. It is difficult for M^3 to provide support to illiterate people because it provides text-based support.

In this study, we have developed a method to provide support to illiterate people engaging in multilingual face-to-face communication.

2 Related Work

In order to provide support to illiterate people, it is necessary to read out the text data verbally. Some studies have employed speech synthesis as a text-to-speech technology [4]. Recently, the quality of speech synthesis has improved. Moreover, the text-to-speech technology has been applied for providing support to visually impaired people [5, 6]. An interface that provided support to the illiterate people has been discussed [7]. On the other hand, the interface that supports the illiterate people is not discussed enough. Both visually impaired people and illiterate people require support by voice data. However, illiterate people require a different type of support

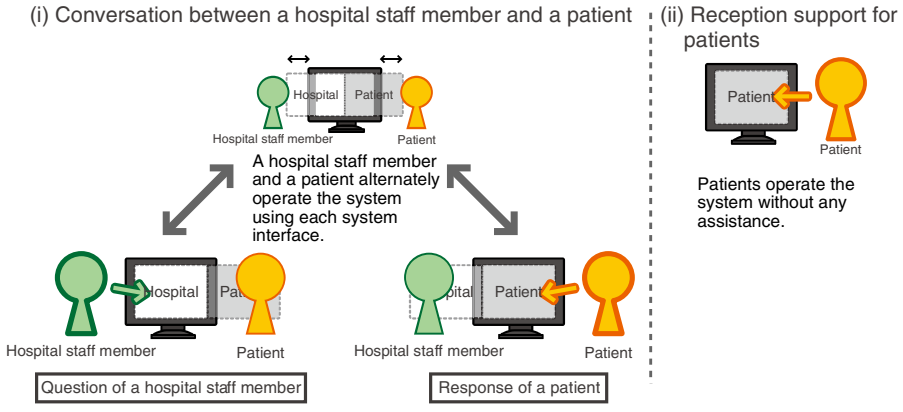

Fig. 1. Image of conversation using M^3

from that required by visually impaired people because the former can look at a display and texts. It is necessary to develop an interface that provides support to illiterate people.

3 M^3: A Support System for Multilingual Medical Reception

1. Summary of M^3

   We have developed a support system for multilingual medical reception termed M^3 [1]. M^3 supports face-to-face communication and the procedure followed at hospital receptions. In order to avoid problems related to translation accuracy, M^3 uses parallel texts that were translated accurately by medical interpreters. M^3 can obtain and share parallel texts using Web services via Language Grid [8]. Users operate M^3 using a touch screen. Figure 1 shows the image of a conversation using M^3. When a hospital staff member and a patient communicate, they alternately operate the system using each system interface. When a patient receives support for following the required procedure at the hospital reception, he/she operates the system without receiving any further assistance. We have installed our system in the Kyoto City Hospital in Japan, and the system has been placed at the reception desk. The system is currently in operation.

2. Problems

   The developed system provides text-based support for conversation and information services. In this system, illiterate people cannot understand the information shown on the display. We received feedback from medical interpreters that the system should provide support to illiterate people. Therefore, we have developed a method for providing support to illiterate people engaging in multilingual face-to-face communication using a touch screen.

4 Text-to-Speech Function in Multilingual Communication Using a Touch Screen

In this study, we have developed a method that provides support to illiterate people when the use the M^3 system (having limitations) that is operated using the touch screen. Although illiterate people cannot understand the meaning of the texts that are shown on the display, they can operate the system by viewing the display. Therefore, illiterate people can use the system if they can understand the meaning of the texts. We provide support to illiterate people by implementing the text-to-speech function. In the system operated by a touch screen, the following problems have to be solved for the realization of the text-to-speech function.

1. Text-to-speech of all texts shown on the display
   We should enable the reading out all texts shown on the display, in order to provide support to illiterate people. Although illiterate people cannot understand the meaning of the texts that are shown on the display, they can operate the system by viewing the display. Therefore, we need to enable a read out of only that text that an illiterate person cannot understand and not all the texts.
2. Screen area
   In the system operated by the touch screen, the sizes of the interfaces operated by users need to be sufficient for the users to touch. The screen area of the system is limited compared to that of a mouse-driven system. Therefore, we require a method in which there is no dependence on the screen area.

We considered that these problems may be solved by the following solutions.

1. Text-to-speech conversion by suitable selections by users.
   If users touch the text shown on a display, only the touched text is read out verbally.
2. Operation of text-to-speech function separate from operation of touch screen.
   The text-to-speech function is not usually used in the system. Therefore, we separate the operation of this function from that of the main functions. The function is physically implemented operated beside the touch screen.

We propose to implement the text-to-speech function using a switch. Texts are displayed on most of the interfaces that are handled by the users of the system. Therefore, we develop two modes of the system: the normal operation mode and the text-to-speech mode. When the text-to-speech mode is in operation, the texts selected by a user are read out verbally. Users can switch between these two modes in order to perform a normal operation or a text-to-speech operation. The selector switch that is placed beside the touch screen is used to switch between the modes. Figure 2 shows the actions by the operation of the selector switch. If users press the selector switch, the clicked text is read out verbally. If users do not press the selector switch, a normal operation is performed.

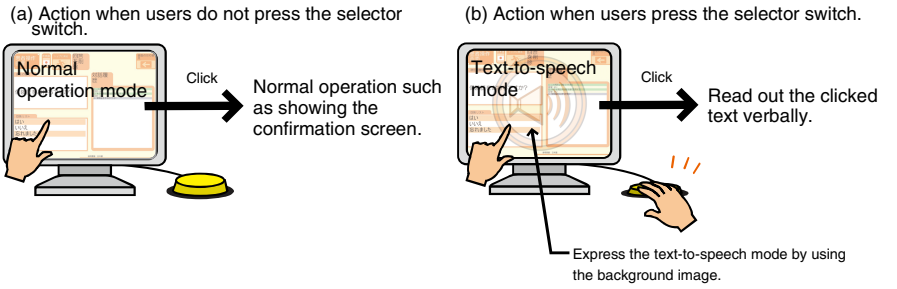

Fig. 2. Actions by the operation of the selector switch

5 Experiment

5.1 Experimental Outline

We performed an information retrieval experiment using the proposed method in order to examine its efficiency and determine the problems encountered in the method. The subjects of the experiment were 12 students from Wakayama University.

In the experiment, the subjects retrieved the answers of five given questions using the Q&A function of M^3. Table 1 shows these five questions.

The subjects performed their tasks under the following experimental conditions.

(a) Texts are readable and subjects retrieve information using the normal operation mode.
(b) Texts are unreadable and subjects retrieve information in both the normal operation mode and the text-to-speech mode.

Each task was performed by six subjects. We compare the results of these experimental conditions.

In the experiment, in order to avoid the influence of the quality of the voice data on the experimental results, we used actual voices that were recorded in advance.

We simulated a situation in which the subjects could not read the text shown on the display, because in reality, the subjects who participated in the experiment were literate. We converted the text shown on the display into unreadable texts in the case of experimental condition (b). Figure 3 shows the screenshots of the experimental tool used for information retrieval. In experimental condition (a), the texts shown on the display are readable. On the other hand, under condition (b), the texts shown on the display are converted into unreadable texts. Under experimental condition (b), if a subject touches an interface with unreadable texts, the texts are converted into readable texts and read out verbally.

After the experiment, we asked the subjects to fill out a questionnaire on the text-to-speech function.

Table 1. Five questions used in the experiment

Question number	Question
1	What is the test/examination charge?
2	How much are parking fees?
3	Where is there a pharmacy?
4	When are medical interpreters available?
5	What are the reception hours and days of the week if you are the previous patient?

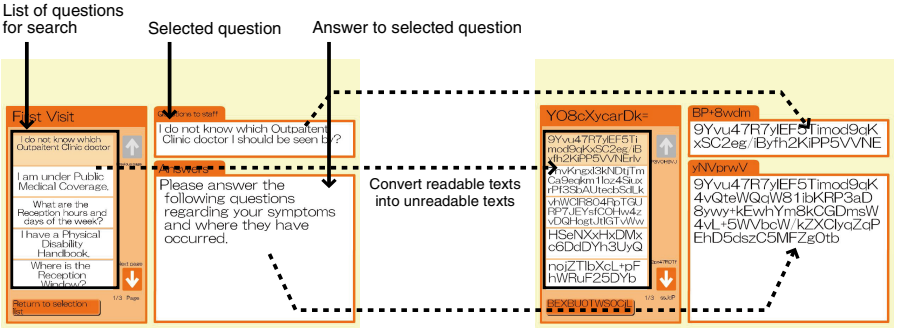

Fig. 3. Screenshots of experimental tool for retrieval

5.2 Results

1. Retrieval time and number of times voice data is played

   Table 2 shows the retrieval time in the experiment. It is the time that the subjects required to search for the answers to the experimental questions. As shown in table 2, the retrieval time using the text-to-speech mode was five times that using the normal operation mode.

   Table 3 shows the number of times voice data of the accurate answers to each question is played. As shown in table 3, the number of times voice data is played was large when the number of letters in the text was large. It is difficult for the subjects to understand the text the first time if the text contains a considerable amount of information. We need to consider this problem in the future.

2. Result of questionnaire

   The evaluation of a five-point scale and a free description of the questionnaire are as follows:

   (a) Evaluation of a five-point scale

   Table 4 shows the results of the questionnaire. We used a five-point Likert scale for the evaluation: 1: strongly disagree, 2: disagree, 3: neutral, 4: agree, and 5: strongly agree.

Table 2. Retrieval time in the experiment

Experimental condition	Subject	Retrieval time (m:s)	Average (m:s)	Standard Deviation (m:s)	Significance probability
Normal operation mode	A	01:18	01:20	00:11	0.004
	B	01:27			
	C	01:18			
	D	01:00			
	E	01:24			
	F	01:33			
Normal operation mode and text-to-speech mode	G	05:07	06:25	01:23	
	H	08:00			
	I	04:23			
	J	07:15			
	K	06:43			
	L	07:04			

Table 3. Number of times voice data of the accurate answers is played

Question number	Number of letters in the answer (letters)	Subject						Average (times)
		G (times)	H (times)	I (times)	J (times)	K (times)	L (times)	
1	36	6	4	4	2	2	2	3.3
2	61	3	5	3	3	6	6	4.3
3	16	1	1	1	2	1	1	1.2
4	45	2	1	1	1	2	2	1.5
5	68	3	5	4	6	8	2	4.7

– Operation of the selector switch

From the results of questions (1), (2), and (3) shown in table 4, we found that the subjects were able to operate the selector switch with ease.

– Support using the selector switch

We proposed the implementation of the text-to-speech function using the selector switch. The selector switch that was placed beside the touch screen was used to switch between the modes. As shown in table 4, the evaluated value of question (4) is 3.5. Therefore, the method using the selector switch has little effect on the operation of the system.

– Effects of using voice data

In the experiment, we used the actual voice data. From the results of questions (5) and (6) shown in table 4, we found that the subjects listened to the actual voice data of a text without any difficulty. However, we received feedback that it was difficult for the subjects to understand all the contents of the text when the text contained a large amount of information. If the text

Table 4. Results of the questionnaire

Questions	Subject						Average
	G	H	I	J	K	L	
(1) I could switch between each mode easily.	5	3	2	4	4	2	3.3
(2) I could play the voice data easily.	5	5	3	2	4	4	3.8
(3) It is difficult for me to touch the screen while pressing the selector switch.	2	1	2	3	2	4	2.3
(4) I think that it is appropriate that the selector switch is used in order to switch between each mode.	5	4	4	4	3	1	3.5
(5) I could understand the contents of the texts by the listening to the voice data.	4	5	4	4	4	5	4.3
(6) I felt that the voice data is hard to catch what it means.	2	1	2	2	1	1	1.5
(7) I think that the voice data helped me when I could not read the text.	4	5	5	2	5	5	4.3

We used a five-point Likert scale for the evaluation: 1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree, and 5: Strongly agree.

contains a large amount of information, people need the intelligible voice data that accentuates important information.

– Providing support to illiterate people using voice data

From the result of question (7) shown in table 4, we found that the use of voice data is necessary to provide support to illiterate people.

(b) Free description of the questionnaire

The merits and demerits that were described by the subjects who undertook the questionnaire are as follows:

– Merits

- The switching between the two modes was very easy.
- I could quickly get used to the operation because the behavior of the selector switch was simple.
- The method to play the voice data was easy to understand.

– Demerits

- I was not able to perform the normal operation while using the text-to-speech mode. I think that the users require a structure that enables them to perform a normal operation while using the text-to-speech mode.
- I think that displaying an icon for playing voice data on a display is more understandable than the proposed method.
- It was exhausting to keep pressing the selector switch.

As a merit, the subjects described that the switching between the two modes was very easy and they could get used to the operation. As a demerit, they described that the proposed method had problems regarding the way to use the selector switch. Therefore, we need to improvise the way in which we use the selector switch.

6 Discussion

From the results of the experiment and the questionnaire, we found that it was very easy for users to switch between the two modes by using the selector switch. However, we also found that by the proposed method, the users faced problems in using the selector switch. The problems with the proposed method are as follows:

1. Users cannot perform a normal operation when they use the text-to-speech mode.
2. Users have to keep pressing the selector switch in order to use the text-to-speech mode.

These problems may be solved in the following ways.

1. Using the text-to-speech mode as an additional mode

In our proposed method, the two modes of the system are independent. Users have to return to the normal operation mode in order to perform a normal operation after the use of the text-to-speech mode. If users use the system for a long time, the number of switching modes will increase.

This problem may be solved by the following method: the text-to-speech mode is used as an additional mode with the main mode. In this method, when the mode of the system is changed to the text-to-speech mode, the icons for playing the voice data are shown on the texts in the display. Texts are read out verbally by touching these icons. If users touch anything except these icons, the system deactivates the text-to-speech mode. This will prevent the increase in the number of switching modes.

2. Switching to the other mode every time users press the selector switch

The main mode of the system is the normal operation mode. Therefore, it is necessary to return to the normal operation mode after using the text-to-speech mode.

We propose the following method as a solution to this problem: the system switches to the other mode every time users press the selector switch. Users have to press the selector switch in order to return to the other mode in this method. There is a possibility that users may forget to return to the other mode. Our proposed method was designed to return to the normal operation mode when users released the selector switch. However, we found that the proposed method imposed a burden on the subjects.

If we apply this solution, we also need to devise a method to prevent users from forgetting to press the selector switch. For this purpose, the system has to return to the normal operation mode automatically if users do not operate the system during a given time. This behavior of the system will enable users to perform a normal operation even if they forget to press the selector switch.

7 Conclusion

We have developed a support system for multilingual medical reception termed M^3. M^3 provides reliable communication between a hospital staff member and a foreign patient using accurate translations called parallel texts. Users operate M^3 using a touch screen and receive text-based support. However, there are foreigners that are Japanese illiterate or are their native language illiterate in Japan. It is difficult for M^3 to provide support to illiterate people because it provides text-based support.

In this study, we developed a method to provide support to illiterate people engaging in multilingual face-to-face communication. We performed an experiment to examine the effect of the proposed method. The results of the experiment are as follows.

1. From the results of the questionnaire, we found that the subjects were able to operate the selector switch with ease. Therefore, we concluded that the method using the selector switch had little effect on the operation of the system.
2. The time for the retrieval of information using the text-to-speech mode was five times that using the normal operation mode. We need to consider a structure that can retrieve the required information easily if many readings of texts are required.

The results of this study can be applied to the development of the support system for illiterate people engaging in multilingual face-to-face communication. Moreover, these results can be used as basic data for the design of a speech dialog system. In the future, we intend to improve the method for providing support to illiterate people on the basis of the results of the experiment.

Acknowledgments. This work was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan.

References

1. Miyabe, M., Fujii, K., Shigenobu, T., Yoshino, T.: Parallel-Text Based Support System for Intercultural Communication at Medical Receptions. In: Ishida, T., R. Fussell, S., T. J. M. Vossen, P. (eds.) IWIC 2007. LNCS, vol. 4568, pp. 182–192. Springer, Heidelberg (2007)
2. Kobayashi, Y.: How to Accept Foreign Patients. Elsevier, Japan (2002) (in Japanese)
3. UNDP, Human Development Report 2007/2008, http://www.undp.org/
4. ElAarag, H., Schindler, L.: A speech recognition and synthesis tool. In: Proceedings of the 44th annual Southeast regional conference, pp. 45–49 (2006)
5. Sporka, A.J., Nemec, V., Slavik, P.: Tangible newspaper for the visually impaired users. In: CHI 2005 extended abstracts on Human factors in computing systems, pp. 1809–1812 (2005)
6. Chirathivat, J., Nakdej, J., Punyabukkana, P., Suchato, A.: Internet explorer smart toolbar for the blind. In: Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology: in conjunction with 1st Tan Tock Seng Hospital Neurorehabilitation Meeting, pp. 195–200 (2007)
7. Leporini, B., Andronico, P., Buzzi, M.: Designing search engine user interfaces for the visually impaired. In: ACM SIGCAPH Computers and the Physically Handicapped, pp. 17–18 (2003)
8. Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration. In: IEEE/IPSJ Symposium on Applications and the Internet (SAINT 2006), pp. 96–100 (2006)