A Multi Modal Supporting Tool for Multi Lingual Communication by Inducing Partner's Reply

*1Kazunori IMOTO, *2Munehiko SASAJIMA, *1Taishi SHIMOMORI, *1Noriko YAMANAKA, *1Makoto YAJIMA and *1Yasuyuki MASAI

*1Multimedia Laboratory RDC Toshiba, 1 Komukai Toshiba Saiwai Kawasaki Kanagawa 212-8582 Japan
*2ISIR, Osaka Univ. Japan, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
{kazunori.imoto, taishi.shimomori, noriko.yamanaka, makoto.yajima, yasuyuki.masai }@toshiba.co.jp
msasa@ei.sanken.osaka-u.ac.jp

ABSTRACT

This paper introduces a new tool for supporting multilingual communication between speakers of different languages. Conventional tools such as electronic dictionaries enable users to communicate basic intentions to others, but are often insufficient to help understand replies. The input of a Japanese sentence in the proposed tool not only produces a translation of the sentences but also displays a window featuring possible answers. The authors have evaluated the function of a prototype system which resulted in a thorough understanding of the merits and comings of the proposed tool.

Categories and Subject Descriptors:

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Prototyping, User-centered design, Voice I/O and Natural Language.*

General Terms: Design, Experimentation, and Languages

Keywords: Speech Processing, Translation, Answer Induction, Retrieving meaning-equivalent sentences

1. INTRODUCTION

In view of the growing numbers of people traveling abroad, the opportunity for contact between speakers of different languages is increasing. Since the acquisition of non-native languages requires considerable time and effort, technology to support communication between speakers of different languages is desired. In communication circumstances in which travelers ask questions to local people whom they encounter for the first time, for example, it is psychologically stressful to make the conversation partner wait. So tools for supporting multilingual communication need to be equipped with a simple interface for input and operation to avoid long interruption of the conversation. Voice input makes it possible to input the user's intention simply and rapidly. Speech translation systems, which are realized by integrating speech recognition, machine translation and text-to-speech synthesis, have already been investigated [1-3]. And some systems have been built on handheld devices [4,5]. However, speech translation systems encounter two problems, namely, recognition error and translation error [6]. Both errors increase users' anxiety, because they may cause troubles in communication. A system to support multilingual communication should resolve or reduce the errors.

In this paper, we define communication as consisting of 1) expression of the speaker's intention, 2) response from the conversation partner, and 3) understanding of the response. By using conventional tools for communication such as electronic dictionaries, speakers can express their intentions. However, if the partner's response exceeds the speaker's comprehension, communication falters. The idea of dual directional speech translation in which both the speaker's and the partner's speech are translated might resolve this point because the user can hear the answer in his/her native language. However, this architecture encounters "recognition error and translation error" in both directions. Also, for such a dual directional speech translation system, it is difficult to expand the number of acceptable languages. Furthermore, it is difficult to explain how to use the device to new users who are speakers of different languages.

To solve these problems, we propose "Global Communicator", a tool for supporting communication between speakers of different languages. Firstly, users can input their intention simply by their voice. We assume Japanese users and the user's voice are recognized by Japanese speech recognition quickly. Secondly, the recognized result is replaced by a similar sentence in database of parallel translation and the replaced sentence is translated based on example-based translation technology. The user can select the most suitable sentence from the candidate list, and the effect of the recognition error is reduced. Furthermore, example-based translation reduces the translation error. Lastly, this tool displays not only the translated result, but also prepared answer candidates to induce the partner to reply by drawing or pointing on the screen. By checking the prepared answer, the users can predict and understand the partner's answer regardless of the ability of aural comprehension of a foreign language. So, communication doesn't falter and the users' anxiety concerning the partner's response is reduced.

In this paper, we discuss the Global Communicator and the results of an evaluation experiment on the prototype.

2. PROPOSED SYSTEM

2.1 System Overview

The system performs speech translation between Japanese and English/Chinese in a travel situation such as shopping and transportation. Figure 2d gives an overview of the translation platform. The system is implemented on a PC-based experimental platform equipped with USB audio input/output devices, a 105mm*80mm LCD display and a touch panel. The user can hold the tool in one hand. The key components of this system are

IUI'06, January 29-February 1, 2006, Sydney, Australia.

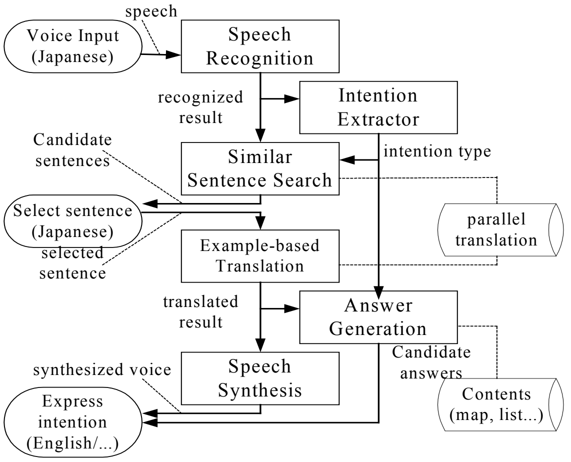


Figure 1. Block diagram of Global Communicator

Japanese speech recognizer, Japanese-English/Chinese translator, English/Chinese speech synthesizer, and Multi-modal user interfaces. Figure 1 shows the processing flow of our translation system. Ellipses, rectangles, and circular cylinders on the figure correspond to the user's operation, process of the tool and the database, respectively.

The Japanese automatic speech recognizer is an HMM-based speaker-independent LVCSR engine using a statistical language model (n-gram). The word dictionary has over 100,000 entries. This recognizer allows users to input their intentions rapidly using their usual words. The translation from the recognized result is carried out using an example-based method. To realize robust example-based translation, example sentences that have meaning-equivalent to the recognized result are retrieved. Retrieval of meaning-equivalent sentences is based on not only word attributes but also the intention type. Retrieved sentences that have higher degree of similarity are displayed on the screen and the most suitable candidate sentence is selected by the user. The selected sentence is translated based on parallel translation corpus. To induce the partner to reply in a manner understandable to the user, the partner's reply is reasoned from the intention type and the attribute of example words. The translated result and induction for reply are read out and displayed in the target language. The method of retrieving meaning-equivalent sentences and reasoning/inducing a reply are discussed in detail below.

2.2 Retrieving meaning-equivalent sentences

The example-based translation method makes the translation result so accurate that the risk of faltering communication is reduced. However, the typical example-based method has the problem that sentences that have different expression from meaning-equivalent example sentences aren't acceptable.

To solve this problem, we propose a method to retrieve the meaning-equivalent sentence. Intention types represent what the speaker want to ask the partner. Table 1 shows the correspondence of the speaker's intention to the intention type.

To explain our proposed method, we pick up a specific example: 1)"Please tell me the place of the abroad buses" as the recognized result, 2) "Where is the bus stop" as an example sentence. At the first step, morphological analysis and dependency parsing are carried out for measuring degree of similarity based on word attribute such as literation, word class or scale reading. In this case, "abroad buses" has high similarity to "bus stop". At the second step, each sentence is categorized as one of the intention types by static rules based on words' attribute and modification relation. In this case, since the expressions "tell me the place" and "Where" both represent the speaker's intention to ask the location, both sentences are categorized as "Where" type. At the final step, the degree of similarity is calculated based on word attribute, modification relation and intention type. As a result, sentences that have higher degree of similarity are retrieved as candidate' sentences.

Table 1. Definition of intention type

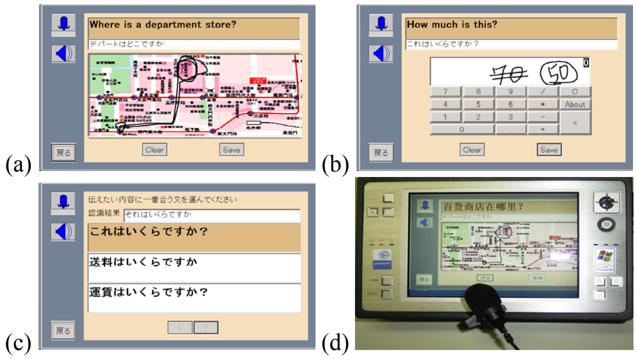
Intention Type	Speaker's intention	Item for understandable answer
	Example sentence	
Where	Ask the location	Map which can be written on
	Where can I find souvenirs?	
Which	Choose alternatives	List of candidates answer
	What color do you have?	
What	Ask definition	Space which can be written on
	What's that?	
How much	Ask Numeric	Calculator which can be written on
	How much is this?	
YN	Confirm Yes/No	Selection screen of Yes or No
	Is there a duty-free shop?	
Request	Relay a request	Space which can be written on
	Please give me a floor guide.	

2.3 Answer Induction

Table1 shows the relation between the speaker's intentions and how to induce the partner's reply. The question to ask the location is categorized as "Where" type. To induce the partner's reply, a map of the neighborhood that can be written on (Figure 2a). By writing the present location or the route to the destination on the map, the user can elicit information about the location without using words. The question to select alternatives is categorized as "Which" type. If the list of candidate answers is countable such as credit card, the list of contents is displayed. By pointing to candidates, the partner can answer quickly and briefly. As to the "YN" question of choosing between affirmation and negation, the Yes/No alternatives are displayed. Regarding the "How much" question for asking the number, the calculator, which can be written on, is display. In price negotiation by means of writing, the user can elicit the number on the screen. Figure2b shows examples of a screen to induce the partner's reply. In the case of "What" type question to ask a specific explanation, such as "What is a popular souvenir from this area?" it is difficult to predict the answer. In this case, a space that can be written on and stored is displayed. Although users can't understand what the partner writes in a non-mother language, by presenting the stored screen to shop assistants at a shop, users can purchase a souvenir.

2.4 User's Operation

The user can operate the tool by touching the button or uttering into a microphone. First of all, the user utters what he/she wants to communicate to the partner in Japanese. Example sentences that have a resemblance to the utterance are retrieved from the database. The top 3 sentences that have higher degree of similarity are displayed on the screen as candidate sentences. The Figure 2c shows the screen for selecting a candidate sentence in the case that the utterance is "How much is this?" By selecting the candidate sentence that is the most similar, the translated result and a prepared answer to induce the partner to reply are displayed. The Figure 2b shows the screen in the case that the



Where is a department store?

デパートはどこですか

How much is this?

これはいくらですか？

伝えたい内容に一番合う文を選んでください

認識結果 それはいくらですか

これはいくらですか？

送料はいくらですか

運賃はいくらですか？

百货商店在哪里？

デパートはどこですか

Figure 2. Screen examples and prototype overview

user selects "How much is it". By bringing the screen to the partner, the partner understands what the user wants to communicate from synthesized speech of the translated result. Scanning through the screen displaying the calculator, the partner is induced to reply using the calculator.

3. EVALUATION

3.1 Procedure of Group Interview

To verify whether the proposed tool can support communication between speakers of different languages, the prototype tool is evaluated using the group interview method. For an experimental trial, 342 sentences, which are often used in shopping/transporting situations, are registered in the database. 22 Japanese and 7 Chinese experimental subjects are selected based on communication ability. The procedure of the group interview is as follows 1) Demonstration and trial use of the tool. 2) Discussion about the problems of the prototype tool, and request of new functions to support communication.

3.2 Results of Experiment

In the group interview, we ask two questions. One is the question for the 22 Japanese subjects about the desire to purchase or rent the proposed tool. 17 subjects respond that they want to purchase or rent the proposed tool. Analysis of subjects' opinions indicates our system is acceptable. The other is the question about the preferred type of answer: "Free Answer" type or "Selecting Answer" type. In the case of the former type, the user and the partner alternately input their intention by speech. In the case of the latter type, communication to induce partners to reply by selecting from predicted answers is supported. According to the results of the questionnaire, the Japanese subjects prefer "Free Answer type". The subjects expressed 14 opinions as to why they prefer "Free Answer type". 10 of them are concerned about the amount of contents and the rest are concerned about circumscribing the partner's answer. On the contrary, the Chinese subjects whose role is the partner prefer "Selecting Answer type", because they can answer quickly and briefly.

4. DISCUSSION

One of the methods to resolve the anxiety concerning the lack of contents is to limit the domain or task. By dividing the situations and collecting the contents according to the situation, the quality of contents become better [7]. However, it is difficult to limit the range of example sentences because of the ambiguity of the travel domain. For example, the shopping situation in travel includes various related conversations concerning such matters as transport to the shops and difficulties encountered during shopping. Even though the situation is restricted to shopping during travel, we don't know exactly how many example sentences should be collected. On the contrary, we think the high quality of contents is guaranteed in specialized applications. From another perspective, this tool is a terminal for extracting information by interview from speakers of a foreign language. With the number of foreigners living in Japan growing every year, opportunities for them to utilize administrative services, hospitals and so on have increased. For example, when treating foreign patients in hospital, doctors who can't speak patient's language can use the proposed tool to elicit information concerning the treatment and explain the therapeutic strategy. Since, unlike in the travel domain, the interview procedure in specialized applications such as hospitals can be formulated and it is not difficult to collect example sentences. In view of the increase in the number of people visiting or residing in communities whose languages they don't speak effectively, difficulties arising from miscommunication can be expected to occur frequently. Moreover, the need for the proposed method of supporting communication can be expected to increase in the future. To enhance practicality, it is necessary not only to limit the task but also to endow the search for similar sentences with greater flexibility [8]. The problem of example-based translation is that the sentences users can utilize are limited to those in the database of example sentences. If degree of similarity between recognition result and example sentence in which a word is replaced can be calculated, flexibility of input becomes higher.

5. CONCLUTIONS

In this paper, we proposed "Global Communicator" to support communication between speakers of foreign languages. The method of supporting communication not only by expressing the user's intention in the partner's language, but also by inducing the partner to reply in an understandable manner was well received by the experimental subjects. However, problems to be solved such as the lack of contents were also clarified. This indicates our proposed system is applicable to at least specialized applications in which the amount of contents is constrained such as hospitals. In the future, we plan to apply our system to the travel domain by solving the lack of contents problems.

6. REFERENCES

[1] A. Lavie, A. Waibel, L. Levin, M. Finke, D. Gates, M.Gavalda, T. Zeppenfeld, P. Zahn. "JANUS III: Speech-to-speech translation in multiple languages", Proc. of ICASSP-97 pp.99-102 (1997).

[2] F.Sugaya, T. Takezawa, A. Yokoo, S. Yamamoto "End-to-end evaluation in ATR-MATRIX: Speech translation system between English and Japanese", Proc. Eurospeech99 pp.2431-2434 (1999)

[3] S. Yamamoto, "Toward Speech Communications beyond Language Barrier -- Research of Spoken Language Translation Technologies at ATR ", Proc of ICSLP2000 vol.4, pp.406-411(2000)

[4] B. Zhou, D. Dechelotte, Y. Gao, "Two-Way Speech-to-Speech Translation on Handheld Devices", Proc of ICSLP2004 pp.1637-1640 (2004).

[5] R. Isotani, K. Yamabana, S. Ando, K. Hanazawa, S.Ishikawa, T.Emori, H. Hattori, A. Okumura ,T.Watanabe,. "An automatic speech translation system on PDAs for travel conversation", Proc.ICMI-02 pp.211-216

[6] Y. Lee, S. Oh, J. Park "Usability Consideration of Speech-to Speech Translation System", Proc. of ICSLP2004.

[7] Y. Wakita, K. Matui, Y. Sagisaka, "Fine keyword clustering using a thesaurus and example sentences for speech translation", Proc of ICSLP2000, vol. 3, pp.390-393 (2000).

[8] O. Furuse, H. Iida, "Constituent Boundary Parsing for Example-based Machine Translation", Proc. of Coling94, pp.105-111 (1994).