7 MASTAR Project

7-1 Speech Translation Technology in MASTAR Project

SUMITA Eiichiro

MASTAR project, Multi-lingual Advanced Speech and Text Research Project was launched on April 2008 at National Institute of Information and Communications Technology. The project includes research and developments aiming to break language barriers between different language speaking people and barriers between human and machines. While researches of language resources, language translation, and speech communication and information analysis are intensively and collaboratively conducted in the project, this paper concentrates, VoiceTra, the application of speech translation technology.

Keywords
Language resources, Translation, Speech communication, Information analysis

1 Background of MASTAR project

Speech and language processing have advanced dramatically in recent years.

This is partly because of establishing the corpus-based technology, i.e., technology of collecting a large volume of data (corpus) and processing the corpus automatically with machine learning algorithms. In conventional approaches, after developing a system, the system is further modified for practical application. In recent methods, on the other hand, data is collected directly when the system is actually used, and the collected data is used for machine learning. Performance can be improved in the research and development phase through the actual utilization of the system. This is a start of new research and development.

The spread of the Web are considered to support further progress. By using the framework of the Web and the information distributed on the Web, one can collect proper nouns from all over the world, develop multilingual dictionaries, collect corpora, look for relationships among words, and analyze the credibility of information.

On this basis, the MASTAR project was planned. The outline of the MASTAR project is as follows.

The aim of the project is to establish a framework of research and development for the continuous collection and growth of speech and language resources, making connections to industries and society through the Web.

The following four areas of research and development are conducted.

(1) Technological development, social experiments, and contributions to society are made for network-based speech translation, which was selected as a Project of Acceleration for Contributions to Society by the Council for Science and Technology Policy.
(2) A machine translation system was developed in collaboration with industry to translate manuals and other documents. A positive growth cycle is established for the

collection of shared dictionaries and shared corpora and for translation technology research.
(3) Technological development of speech interface technologies is made as part of universal communication.
(4) World-wide language resources are developed and distributed.

In this paper we focus on VoiceTra, a speech translation application. Related technologies are explained in "**3-1** Overview of Spoken Language Communication Technologies", "**3-3** Multilingual Speech Synthesis System", "**4-1** Special-Purpose System for Multi-Lingual High-Quality Translation", and "**5-1** Information Analysis Technology at NICT".

2 Project for the "Realization of Speech Communication Technology overcoming Language Barriers"

This project, which we explain here, was selected as one of the Projects for the Acceleration of Contributions to Society, by the Cabinet Office, in April 2008.

Language barriers are an important issue in the borderless society. For example, the New Growth Strategy[*1] aims "to increase the number of foreign visitors to 25 million, three times as many as current visitors, by 2020", which is expected to produce economic effects of 10 trillion yen and new employment of 560,000. However, foreign visitors are not satisfied with the use of foreign languages in public transportation, accommodation facilities, and restaurants.

The goal of the project, as part of the Project for the Acceleration of Contributions to Society, by the Cabinet Office, is "to plan and promote experiments in the fields of travel and shopping, taking account of the present needs and the expected technological improvement in the next five years for automatic speech translation systems that allow communications with foreign people through direct conversations, overcoming the language barrier in the course of internationalization, an starting business services in industry shortly after the project is finished in order to accelerate contributions to society." To be more specific, a network-based speech translation technology is to be established by using translation knowledge distributed over the Web and combining it with translation terminals, in order to collect place names and other proper nouns and handle a variety of text content. Also, innovations using speech translation communication technology are to be made clearly "visible" for smooth commercialization and promotion of the system. For this "visualization", field experiments are to be conducted repeatedly through the collaboration of developers and users, in order to accelerate the contribution of speech translation technologies to society.

3 Achievement of network-based speech translation project: VoiceTra

3.1 VoiceTra

VoiceTra[*2] is automatic translation software that works when users speak to their smart phones. The following are the details of this software. Using this software that can be downloaded without charge, Japanese speakers and English speakers can converse with each other as shown in Figs. 1 and 2. When the speaker puts the phone to their ear, it vibrates for a short time. The speaker then speaks to the phone and translated speech comes back. The first window in Fig. 1 shows the result of speech recognition by the system, and the translation result is given in the third window. The second window shows the result of "reverse translation" (translation of the translated text back to source language), from which one can see if the translation is appropriate. VoiceTra was released in August 2010 and had been downloaded 600,000 times as of

＊1 http://www.kantei.go.jp/jp/sinseichousenryaku/
＊2 http://mastar.jp/translation/index.html

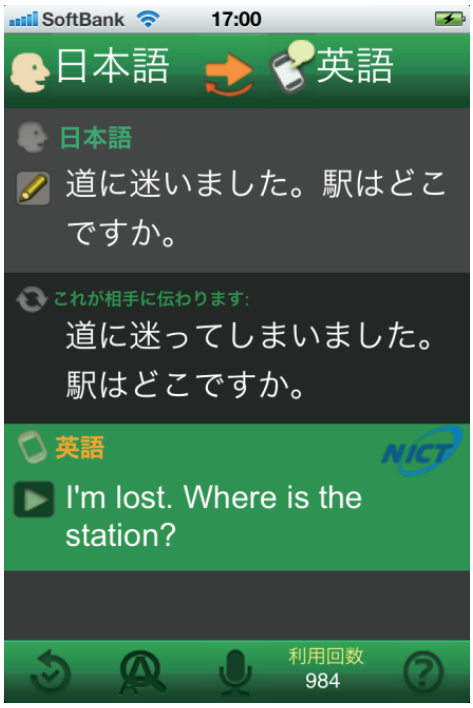

Fig.1 *Screen of VoiceTra (for question)*

Fig.2 *Screen of VoiceTra (for answer)*

March 2012. One out of 200 Japanese people use VoiceTra. This fact indicates that many people now know of this speech translation technology. As will be mentioned below, in collaboration with private companies, this is to be one of NICT's flagships in terms of contributions to society.

3.2 "Basic" speech translation technology

Figure 3 shows an example where Japanese speech is recognized to make Japanese text and the text is translated into English text, which is then synthesized to English speech. In the speech recognition module, the input speech is converted to a phoneme sequence by comparing the speech with an acoustic model that is formed (for each phoneme) from the speech data of many speakers. Then the phoneme sequence is converted to maximize the word sequence probability (which is given by the so-called language model). In this conversion, appropriate word sequences are obtained based on the occurrence probabilities of trigram word sequences, which are acquired from a large volume of Japanese texts. The translation module then selects appropriate English words that correspond to each of the Japanese word sequences and determines the order of the English words. For the selection of English words in accordance with the Japanese word sequence, a translation model that is obtained from Japanese-English parallel texts is used. For the determination of the order of English words, a large volume of English texts are used. Appropriate English word sequences are obtained by using the occurrence probabilities of trigram word sequences. Then the word sequence is sent to the speech synthesis unit, which determines the English word sequence with the intonation. Then it creates waveforms that fit to the intonation, in a certain time unit, called a frame, to suit the speech features that are obtained from a large amount of speech data. The waveforms are finally combined to synthesize a voice.

Since the system is developed automatical-

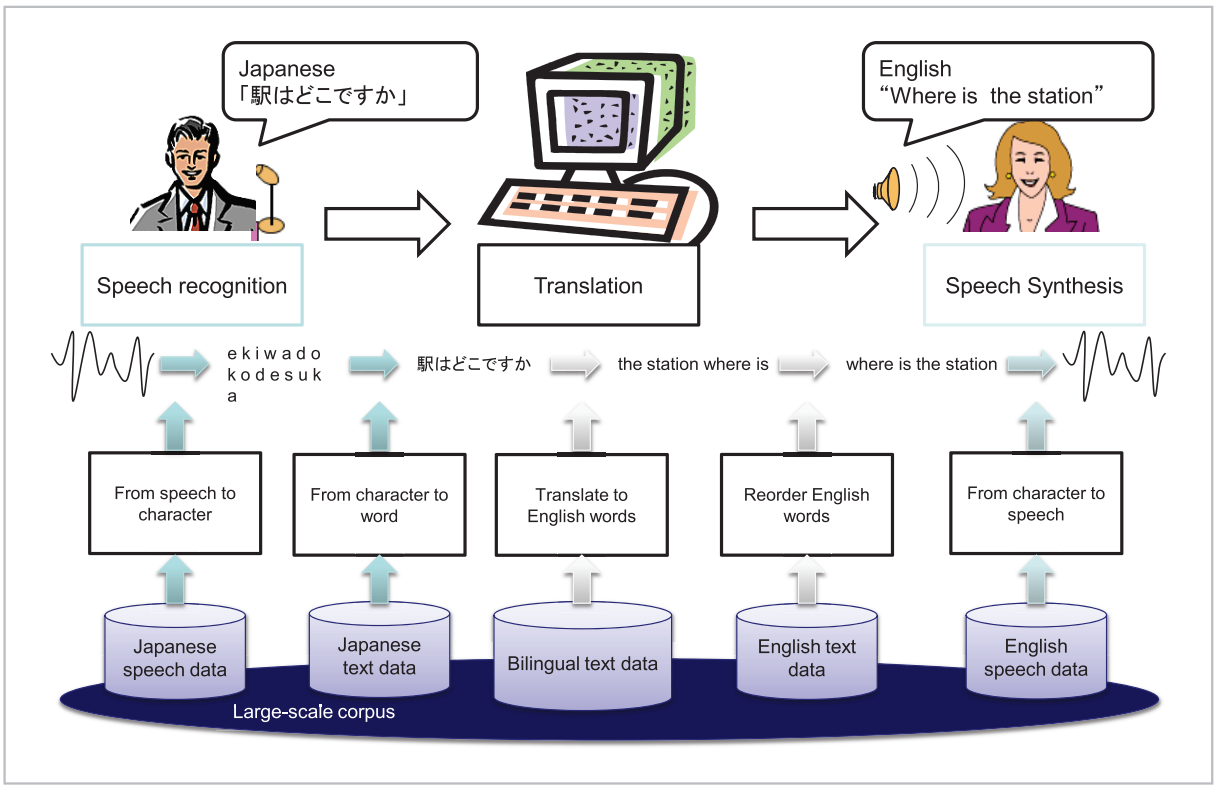

Fig.3 *Basic technology of speech translation*

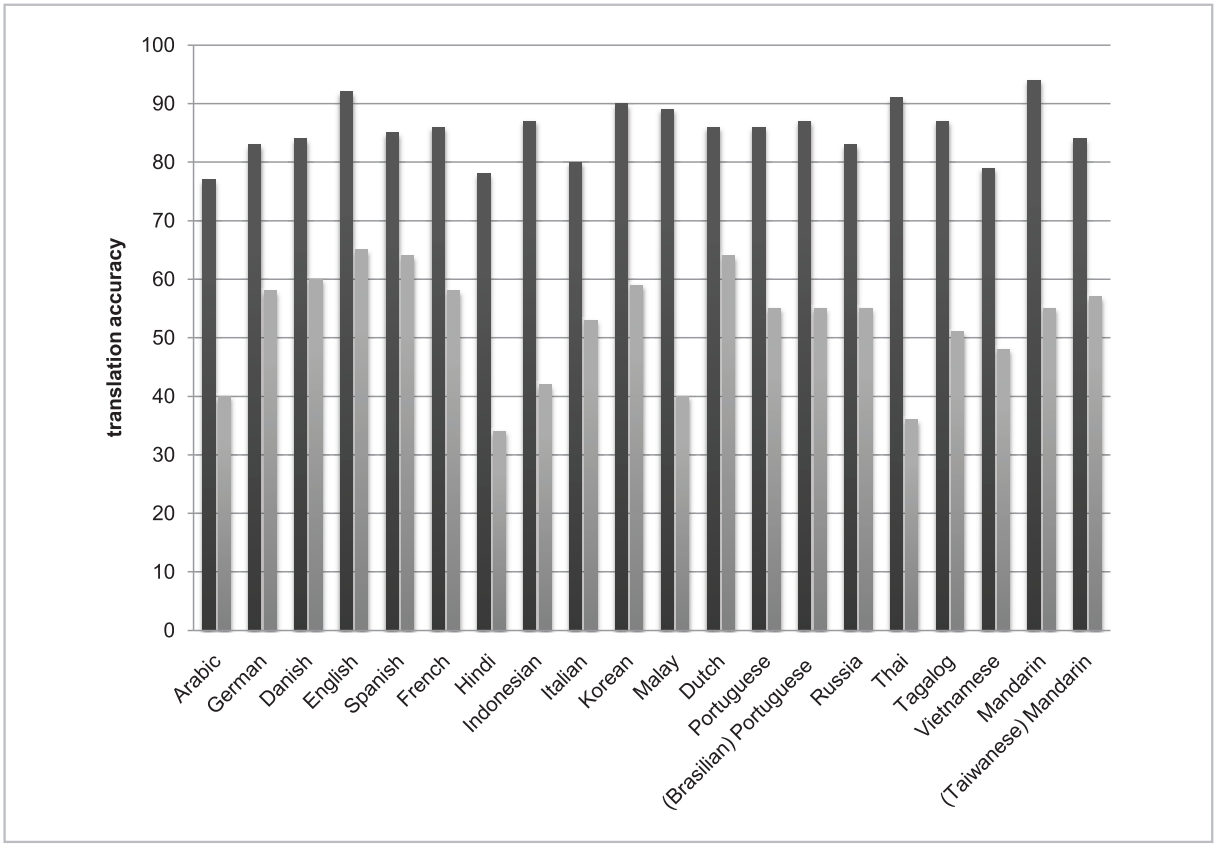

Fig.4 *Comparison of translation probability (Vertical axis: Translation probability to Japanese; Horizontal axis: Source language)*

ly on the basis of a large-scale corpus (Japanese word data, parallel texts, English speech data, etc.) shown at the bottom of the figure, it is called corpus-based technology.

3.3 Performance of translation software

The target of VoiceTra is travel conversation. VoiceTra's translation performance is almost equivalent to a person who scores 600 on TOEIC. The characteristic of VoiceTra is that it is not only multilingual but also high quality. The figure in Fig. 4 shows a comparison between the translation probability (translator-evaluated probability of making an appropriate translation) of conventional multilingual software (light gray) and that of NICT's software (dark gray).

3.4 Network-based speech translation technology and its realization

The basic technology in **3.2** is then applied to a wireless network-based system. Its portability, which cannot be achieved by a system confined to a single computer, the expandability of languages and vocabulary, and automated improvement of performance could accelerate the practical realization of the system. (1) Since it weighs only about 100 g, the client system is highly portable and practical. (2) Since the server has no limitations to its hardware, the languages and vocabulary can be considerably expanded and field data can be used to improve the performance autonomously. In fact, when we used part of the data of VoiceTra for the improvement of speech recognition, precision was improved by 5–10%, although this depended on the target language.

4 Field experiments

Speech translation technology cannot be commercialized even if it has a distinguished base performance such as halved processing time or halved error rates. Also, the error rate of speech translation cannot be decreased to zero, just as that of other pattern recognition technologies cannot. It is therefore necessary that users actually use the system and that the utilization results are fed back by questionnaire surveys or logs to improve and commercialize the system. Field experiments need to be conducted for this purpose. The details of these field experiments are described in "**7-2** Speech-to-Speech Translation System Field Experiments in All Over Japan" and "**7-3** VoiceTra Field Experiments".

Speech translation technology cannot be made multilingual by a single organization because of the high cost. To solve this problem, it is necessary to standardize the protocol needed for speech translation systems and to collaborate between various organizations all over the world. The details of this standardization and collaboration are described in "**7-4** International Standardization of Network-Based Speech-to-Speech Translation Technologies and Expansion of the Standardization Technologies by the International Research Collaborations".

5 Commercialization

NICT and the Narita International Airport Corporation (NAA) conducted field experiments toward the commercialization of the speech translation system in the period from October 4, 2010 to February 25, 2011. 1,600 proper nouns related to the airport (including airline names, tourist spot names, station names, and product names) were added to a glossary. This achieved correct speech recognition and translation. For example, incorrectly recognized results due to insufficient glossary, "ana no kaunta ha doko desuka (Where is the hole's counter?)," is now recognized correctly as "ANA no kaunta wa doko desuka (Where is the ANA counter?)." NAA decided to use the network-based speech translation technology as a solution for foreigners to overcome language barriers. At the end of December 2011, NAA started a speech translation service, with which travelers can download an application (called NariTra; Fig. 5) to their smart phones.

So far there are five transfers of VoiceTra technology, including this case.

NICT conducted a five-year project, "Realization of Speech Communication

Technology to Overcome Language Barriers," which is one of the Projects for the Acceleration of Contributions to Society. The project was supposed to finish at the end of fiscal 2012. However since NICT could obtain significant achievements, more than expected in project planning, the project finished successfully at the end of fiscal 2011, a year earlier than planned. Only this project, out of six Projects for the Acceleration of Contributions to Society, had such success, and it was viewed very positively. The following is a quotation from Section 3.3 "Future plan" in the conference materials from the meeting of ministers and intellectuals at the Council for Science and Technology Policy of the Cabinet Office. "Considered in a comprehensive manner, we determined that the project has achieved its final target of accelerating the realization of an environment where ordinary travelers can enjoy foreign trips in Japanese, English, or Chinese without feeling language barriers. We therefore consider it appropriate that the project was finished at the end of fiscal 2011, a year earlier than the planned end of the project, i.e. end of fiscal 2012."

Fig.5 Poster of NariTra

6 Future of speech translation research

Basic research into speech translation technology was started in 1986, and VoiceTra is a milestone indicating that the technology has finally been commercialized.

Current speech translation technology still has some problems. For example it cannot handle long sentences, and it cannot understand the context of text. NICT is now working on these difficult research targets and aims to realize its next big dream, i.e. simultaneous interpretation of news programs and meetings.

Acknowledgments

For the field experiments which were essential for the present study, we received support from the following parties. We would like to express sincere thanks to them.

- Volunteers who joined the field experiments
- Organizations and corporations which performed the field experiments

Docon Co., Ltd.
NEC Software Hokkaido, Ltd.
HOKKAIDO TOURISM ORGANIZATION
Japan Tourism Marketing Co.
JTB Hokkaido Corp.
Furano Tourism Authority
Furano City Office
Shiretoko Shari-cho Tourist Association
RADIO FURANO
Tabi-no-Yado STELLA, Kamifurano
Country Inn Furusatoso
JTB Global Marketing & Travel Inc.
NEC Corporation
Dr. Hitoshi Yoshida (Department of International Studies and Communications, Faculty of Global Policy Management and Communications, Yamanashi Prefectural University)
Hisashi Yano (Yamanashi Prefecture)
Kouji Ishida (Yamanashi Tourism Organization)

Ekitan & Co., Ltd.
JTB Central Japan Corp.
ATR-Trek Co., Ltd
Kansai Economic Federation
City of Osaka
Osaka Prefectural Government
Osaka Convention & Tourism Bureau
Nara Prefecture
City of Nara
Nara City Sightseeing Information Center
All Nippon Airways Co., Ltd.
Hotel New Hankyu
SHANGHAI FERRY CO., LTD.
NATIONAL STUDENTS INFORMATION CENTER
Kumamoto Institute of Computer Software, Inc.
Kyushu Industrial Transportation Holdings Co., Ltd.
JTB Kyushu Corp.
Not-for-Profit Organization, Industry-academic-government, Nagasaki
KYOTO INDUSTRIAL SUPPORT ORGANIZATION 21
Advanced Telecommunications Research Institute International (ATR)
TOEI KYOTO STUDIO Co., Ltd.
WILLCOM Inc.
SILK LABORATORY Corporation
INTAGE Inc.
UQ Communications Inc.
Yasaka Limousine & Taxi Service corp.
City of Kobe
Foundation for Kobe International Medical Alliance (Kobe IMA)
Mizuho Information & Research Institute, Inc
Tokushima Prefectural Government (Commerce, Industry & Lobour Department)
Tokushima University Hospital
Kyoto Prefecture
KUMAMOTO KENRITSU KUMAMOTO ROHGAKKOH (Kumamoto School for the Deaf)

●Field experiments for commercialization
Oriental Land Co., Ltd.
NARITA INTERNATIONAL AIRPORT CORPORATION

●Commercialization
NARITA INTERNATIONAL AIRPORT CORPORATION
Feat Inc.
ATR-Trek Co., Ltd

●ITU-T
Yushi Naito (Mitsubishi Electric Corporation)
Yoshinori Gotoh (NTT)
Prof. Sakae Okubo (Waseda University)
Andrea Saks

●U-STAR field experiments
Dr. Seung Ku HWANG (Senior Vice President)
Pansak SIRIRUCHATAPONG, D. Ing. (Executive Director)
Dr. Ir. Marzan A. ISKANDAR (Chairman)
Prof. Bo XU (Deputy President)
Dr. George VARKEY (Executive Director)
Assoc. Prof. Thai Quang VINH (Director)
Prof. Kin Mun LYE (Acting Executive Director)
Phuntsho TOBGAY (Director)
Dr. Waqar MAHMOOD (Director)
Dr. Sanat Kumar BISTA (Chief Executive Officer)
Dr. Damsinsuren BAYANDUUREN (President)
Dr. Tumur-Ochir SANJBEGZ (President)
Dr. Gihan WIKRAMANAYAKE (Director)
Dr. Sergio S. CAO (Chancellor of UP Diliman)
Mrs. Michele SAUMON (Director, CNRS)
Dr. Leonel Sousa (President of INESC-ID Lisboa)

●Persons involved in “Realization of Voice Communication Technology to Overcome Language Barriers,” a Project of Acceleration of Contributions to Society
Naoki Okumura, Executive Member, Council for Science and Technology Policy
Shinichi Shimizu, Specially Appointed Professor, College of Tourism, Department

of Tourism and Hospitality Management, Rikkyo University
Junichi Tsujii, Principal Reseacher, Microsoft Research Asia
Prof. Yoshinori Sagisaka, Graduate School of Global Information and Telecommunication Studies, WASEDA University
Kiyoshi Hatashimo, Industrial Science and Technology Policy Division, Industrial Science and Technology Policy and Environment Bureau, Ministry of Economy, Trade and Industry
Takahisa Kashiwagi, Director, Japan Tourism Agency, International Tourism Policy Division
Agency for Cultural Affairs
Prof. Satoshi Nakamura, Graduate School of Information Science, Nara Institute of Science and Technology (NAIST)
Minoru Eto, Managing Director of Service & Solution Development Department, R&D Center, NTT DOCOMO, Inc.
Fumio Watanabe, Executive Director and General Manager, Technology Sector, R & D Strategy Division, KDDI CORPORATION
Mitsugu Miura, Manager, NEC Corporation
Shinta Kimura, Senior Vice President, ANIMO Limited
Akira Ushioda, Research Fellow, Media Processing Systems Laboratories, Fujitsu Laboratories Ltd.
Naoki Kajita, Deputy Director-General, Cabinet Office
Yoshihiro Oishi, Deputy Director-General, Cabinet Office
Shunya Itami, Counselor to Director General (for Science and Technology Policy, Innovation), Cabinet Office (for common fundamental technology (ICT))
Kiyonori Ohtake, Counselor to Director General (for Science and Technology Policy, Innovation), Cabinet Office (for common fundamental technology (ICT))

References

1 KASHIOKA Hideki, "Overview of Spoken Language Communication Technologies," Special issue of this NICT Journal, 3-1, 2012.
2 SHIGA Yoshinori, "Multilingual Speech Synthesis System," Special issue of this NICT Journal, 3-3, 2012.
3 SUMITA Eiichiro, "Special-Purpose System for Multi-lingual High-Quality Translation," Special issue of this NICT Journal, 4-1, 2012.
4 YASUDA Keiji and MATSUDA Shigeki, "Speech-to-Speech Translation System Field Experiments in All Over Japan," Special issue of this NICT Journal, 7-2, 2012.
5 MATSUDA Shigeki, YASUDA Keiji, and KAWAI Hisashi, "VoiceTra Field Experiments," Special issue of this NICT Journal, 7-3, 2012.
6 HORI Chiori, "International Standardization of the Network-Based Speech-to-Speech Translation Technologies and Expansion of the Standardization Technologies by the International Research Collaborations," Special issue of this NICT Journal, 7-4, 2012.

(Accepted June 14, 2012)

***SUMITA Eiichiro**, Dr. Eng.*
Director, Multilingual Translation Laboratory, Universal Communication Research Institute
Natural Language Processing, Machine Translation