Data utilization and analytics platform
Advanced Speech Translation Research and Development Promotion Center

Director General　Yutaka Kidawara

The Advanced Speech Translation Research and Development Promotion Center (ASTREC) promotes research and development of multilingual speech translation technology and its social implementation. Our work is based on Japan's Global Communication Plan (GCP), which aims to eliminate the world's language barriers and facilitate human interaction on a global scale, while forming part of a nationwide initiative that includes skilled researchers and engineers both from NICT and private companies. We aim to accelerate open innovation using multilingual speech translation technology to realize an advanced ICT-based society where language barriers do not exist. In FY2017, we continued to make efforts to reduce the language barriers faced by foreigners visiting Japan for the Tokyo 2020 Olympic and Paralympic Games by improving the accuracy of our multilingual speech translation technology and expanding the range of languages and fields in which it can operate. We also reflected these capabilities in our multilingual speech translation app VoiceTra, and conducted field experiments in collaboration with organizations and private companies from various fields such as disaster prevention, rail travel, shopping, taxi services, medicine, emergency and rescue, and policing. Some of these experiments have yielded new commercial services.

R&D of multilingual speech recognition technology

As the basis of our speech recognition technology, we built a speech corpus consisting of a total of 2,265 hours of recorded speech: 500 hours of Korean, 542 hours of Thai, and 516 hours of Myanmar, To improve the accuracy of speech translation in fields related to travel and daily life, we increased the size of the Japanese-English bilingual dictionary from 100,000 words to 300,000 words, and we also increased the size of the Japanese-Chinese and Japanese-Korean dictionaries from 100,000 words to 210,000 words, respectively. We added 60,000 new words of translation for Thai, Vietnamese, Indonesian, Myanmar, Spanish, and French, respectively. The improvements made to our speech recognition models significantly increased the recognition accuracy for Japanese, Thai, Vietnamese, Indonesian, and Myanmar, with a reduction of between 28% and 42% in word error rates. These improved models have been incorporated into the VoiceTra field trial system and have been made available to the public.

R&D of multilingual speech synthesis technology

To improve the practicality of our Korean and Vietnamese speech synthesis systems, we increased the scale of the speech corpus used to train the acoustic model for each language to 15,000–20,000 utterances (15–20 hours) for both male and female speakers, corresponding to 2–5 times the size of the original corpus. This resulted in a highly accurate acoustic model and better speech synthesis quality. We also improved the pronunciation accuracy of each language by introducing a new text normalization process that transforms non-phonetic characters like numerals and symbols into strings of phonetic characters that are more suitable for reading. The new and improved speech synthesis system has been incorporated into VoiceTra and made available to the public.

As in the speech recognition field, deep learning approaches have also been introduced to the speech synthesis field in recent years, and have resulted in a higher quality of synthesized speech compared with conventional methods based on hidden Markov models (HMMs). At ASTREC, we have been conducting research on deep learning since 2015 and have developed a new speech synthesis system that utilizes deep neural networks (DNNs). Figure 1 compares the new system with a conventional HMM system, and Fig.2 shows the results of speech synthesis listening tests performed using a DNN acoustic model of a Japanese female speaker that we developed. The speech quality of the DNN system was clearly better, achieving an average opinion score 0.6 points higher than that of the conventional system. The Japanese female voice DNN synthesis system has been made publicly available on VoiceTra.

R&D of machine translation technology

Our translation corpus of spoken language in ten different languages for multiple fields including medicine has been expanded far beyond the original target of one million sentences. By using this translation corpus, we have confirmed that our translation system has made steady improvements in accuracy for all languages. In this way, we were able to surpass our original goal in building the translation corpus (which was to provide the foundations needed for the development of multilingual translation services), and made a significant contribution to accelerating the implementation of speech translation technology in society.

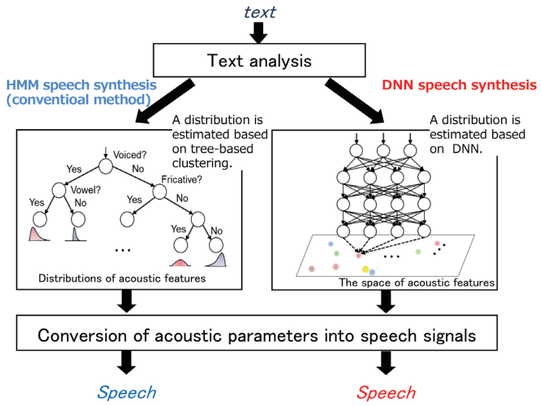

Fig.1 : The outlines of speech synthesis methods

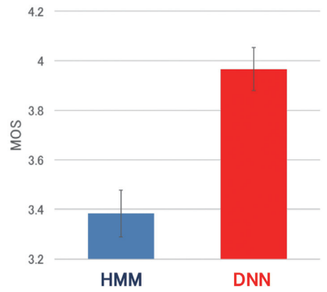

Fig.2 : Results of perceptual evaluation of synthetic speech based on MOS (Mean Opinion Score)

In cooperation with the Ministry of Internal Affairs and Communications, we have also introduced a "Hon'yaku (Translation) Bank" scheme whereby NICT can collect translation corpora from diverse sources scattered throughout Japan. This scheme uses various means (including uploads via the Web) to efficiently collect multilingual translations in multiple fields from around the country, and is expected to achieve greater precision in general-purpose machine translation (Fig.3). This Hon'yaku Bank scheme can be described as a new donation-based collection method.

R&D of a simultaneous language interpretation platform

We have developed a prototype one-shot speech translation system as the first stage of our research and development aimed for completion in 2020. In its current system, VoiceTra must send a series of four requests to the server in order to process one translation—one to perform speech recognition, another to perform text translation, one more to perform reverse translation, and a final request to perform speech synthesis. With a one-shot speech translation system, the input speech is sent to the server as a speech translation request, the server performs all the necessary processing during this one single request, and then sends back the speech recognition, translation, reverse translation, and speech synthesis results in sequence as a single response. This results in a faster response speed than the current system where four round trips of requests and responses are required. In a comparative evaluation we performed in Europe, the current method had a reaction speed of about 6 seconds, while the one-shot speech translation method had a reaction speed of about 2 seconds. In Japan, the reaction speed increased from about 2 seconds to about 1.5 seconds. In the future, we will incorporate the one-shot speech translation method into VoiceTra. We also plan to expand this system to work with continuous speech input, which would enable the development of a system that can perform simultaneous interpretation of lectures and the like, and to provide a research platform for simultaneous interpretation.

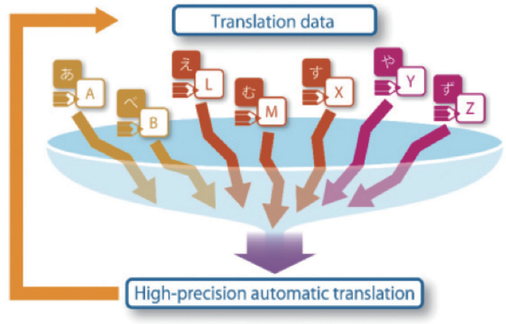

Fig.3 : The "Hon'yaku Bank" concept

NICT team won awards at the World Robot Summit 2018

The NICT Team formed by Dr. Komei Sugiura, Dr. Aly Magassouba, and others from the Advanced Speech Technology Laboratory of ASTREC, won the "1st Place (METI*1 Minister's Award)" and the "JSAI*2 Award" in the Partner Robot Challenge Virtual Space at the World Robot Summit (WRS) 2018. WRS 2018 was held from Oct. 17 - 21 at Tokyo Big Sight and was sponsored by METI and NEDO*3 . In this challenge, seven domestic/international teams competed with each other and they were judged based on the achievement rates of the following 3 tasks: 1) multimodal language understanding task—how accurate can the service robot in virtual space understand users' commands using non-linguistic information such as images; 2) gesture recognition task in the same virtual environment; and 3) multimodal language generation task. The NICT Team marked the best achievement rates in all 3 tasks leading themselves to their victory.

The NICT Team being awarded the "METI Minister's Award" by Parliamentary Vice-Minister of Economy, Trade and Industry, Mr. Akimasa Ishikawa

*1 Ministry of Economy, Trade and Industry
*2 The Japanese Society for Artificial Intelligence
*3 New Energy and Industrial Technology Development Organization