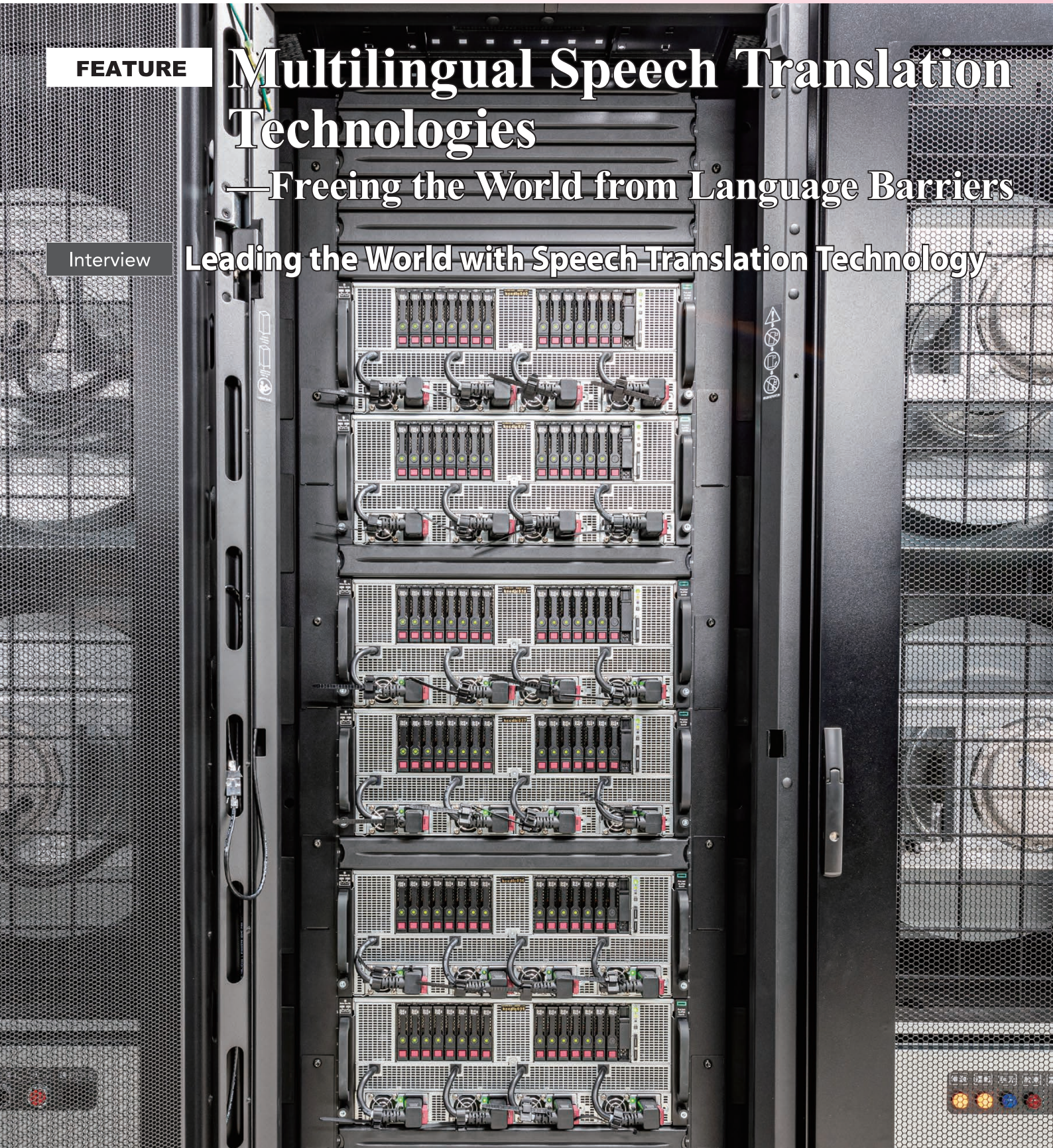
NICT NEWS
National Institute of Information and Communications Technology
No.2
2020
Vol.480



FEATURE

Multilingual Speech Translation Technologies
—Freeing the World from Language Barriers

Interview

Leading the World with Speech Translation Technology

CONTENTS

NICT NEWS 2020 No.2 Vol.480

National Institute of
Information and Communications Technology

FEATURE

Multilingual Speech Translation Technologies —Freeing the World from Language Barriers

Interview

TOPICS

Cover photo: Computer Room
Using machine language methodology such as deep learning based on the massive accumulation of corpus, or a language database, large-scale computation is conducted in this room to build models for speech processing including speech recognition, machine translation, and speech synthesis, and for natural language processing.

Upper-left photo:
Front view of the Universal Communication Research Institute, one of the NICT Kansai bases, located in Keihanna Science City.
This building houses the Advanced Speech Technology Laboratory, Advanced Speech Technology Laboratory, System Development Office, and Planning Office of the ASTREC.

Interview

Leading the World with Speech Translation Technology

Advanced Speech Translation Research and Development Promotion Center (ASTREC)

KIDAWARA Yutaka

Director General
Advanced Speech Translation Research and Development Promotion Center (ASTREC)

He joined the Communications Research Laboratories (currently NICT) in 2001 and has been engaged in R&D on technologies including multilingual speech translation, speech processing, information analysis, information services, and universal communication based on ultra-realistic presence technologies. Ph.D. (Engineering).

Automatic language interpretation by machine was merely a dream several years ago. Now, translation apps for smartphones and AI translators have started to spread quickly.

Translating foreign languages used to be difficult for a machine but why has performance improved so dramatically in recent years? We talked with Dr. KIDAWARA Yutaka, Director General of Advanced Speech Translation Research and Development Promotion Center (ASTREC), NICT, the leader in machine translation and speech recognition/synthesis technologies in Japan.

Advanced speech translation technology

—— What is the advanced speech translation technology of ASTREC?

KIDAWARA ASTREC is researching and developing multilingual speech translation technologies, which consist of three elemental technologies: speech recognition, translation, and speech synthesis.

First, an input speech is converted into text by multilingual speech recognition technology, which is then translated into the target language by multilingual translation technology. Finally, the translated text is read out by multilingual speech synthesis technology. All three of these technologies are being studied here at ASTREC. We are also working to turn the results of our development into commercial software packages which we then release to the public.

——And one of the fruits of those efforts is VoiceTra?

KIDAWARA That's right. It's a free smartphone app that supports translation between 31 languages, 18 of which support speech input and speech synthesis is available in 16 languages.

——The supported languages include many Asian languages.

KIDAWARA Asian countries are important to Japan economically and politically, and we have many tourists from Asia. That's why we focus on these languages. Another reason is to differentiate our system from Google Translate which has the edge in European languages.

Since 2014, the Ministry of Internal Affairs and Communications (MIC) has been promoting the "Global Communication Plan" under the slogan of "eliminating language barriers from the world." We have been accelerating our R&D to make this slogan a reality, especially for the many that will visit Japan for the Tokyo Olympics and Paralympics which will be held this year. VoiceTra is currently designed for daily and travel-related conversations, but we plan to improve our system to support foreign nationals who stay for longer periods in Japan overcome the language barriers that they face in their day-to-day lives.

Outcomes of machine translation using neural networks

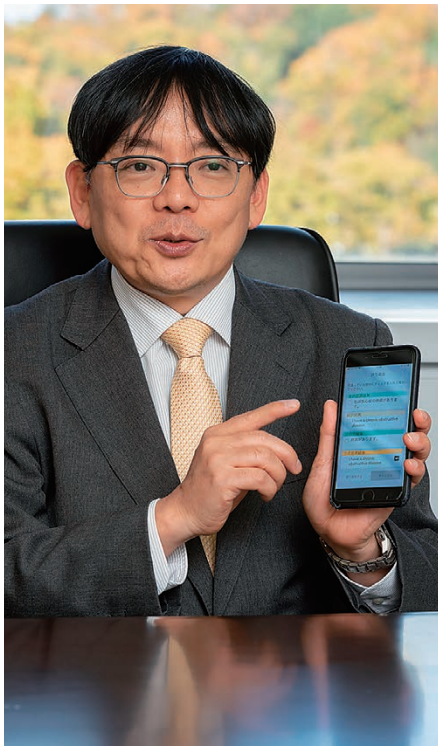
——What is machine translation technology?

KIDAWARA Machine translation is, in its literal sense, a technology for translating languages by machine. Today's machine translation uses neural networks, which is capable of collecting a large amount of parallel corpus (a set of original and translated texts) and refines the system by deep learning. The basic framework is developed by hand, but errors reported by the users are also used to improve the translation accuracy.

One of our latest accomplishments is the implementation of the automatic language

Interview

Leading the World with Speech Translation Technology



identification feature to VoiceTra. This allows speakers of different languages to start a conversation without having to configure the language settings. It identifies what language is being spoken from the first 1.5 seconds of utterance and has achieved around 90% in accuracy. The language identification technology comes in handy for call centers, where they receive random calls in different languages and often have faced the issue of not understanding what language the caller is speaking which has kept them from using translation systems that requires language settings in advance. The technology can be useful for "AI speakers" which are gaining popularity, where users are also random.

A pocket-sized translator for daily and travel conversations, "Pocketalk," is another application example of our research results, and it employs the translation engines developed by NICT.

——Are there any other ongoing technical developments concerning voice translation?

KIDAWARA We have also developed a technology for isolating the human voice from noise. Recent enhancements in deep learning capability have shown that learning a large amount of data containing noise is more effective to improve recognition accuracy than focusing on separating noise and voice.

Another research focus is recognizing speech of a distant speaker. The VoiceTra app for smartphones assume that the user speaks close to the microphone, but devices such as AI speakers must accurately recognize utterances of speakers in various distances which is often the case in real situations. The technology for distant speech recognition is now being studied.

We also need to focus our R&D target on technologies for identifying the voice of a specific person in a crowd and recognizing that person's speech.

——Synthesized speech is also impressively realistic.

KIDAWARA Speech synthesis technology is developing rapidly and has now reached the level where it is hard to distinguish whether the voice is synthesized or of an actual human. In fact, the telephone answering machine at the Keihanna Office uses a synthesized voice, but thanks to the efforts of our researchers, you may not realize it's synthesized and not pre-recorded.

I think it would be interesting if VoiceTra could read out the translated text while mimicking the voice of the original speaker. It may not be a distant dream to create a system that analyzes the speaker's voice and then speaks in a different language but in the same voice.

——Speech translation technology has made rapid progress recently. What was the breakthrough?

KIDAWARA After 30 years of development, speech translation has finally become practical. During this period, the development experienced two paradigm shifts. Initially, sentences were analyzed based on rules, or syntax. However, the number of rules that can be defined manually was limited and restricted the accuracy of translation. In around 2000, statistical translation technology emerged, which probabilistically calculates the word order using a parallel corpus. This was the first paradigm shift which allowed the technology to almost reach the practical stage but not quite for full-fledged application.

Then, in around 2016, translation technology using deep learning with a neural network was introduced by Google. This was the second paradigm shift. Rule-based translation uses more than 10,000 rules, statistical translation uses hundreds of thousands of parallel translation sentences, and neural translation requires several million parallel translation sentences.

The use of large-scale data with neural networks not only enhanced the translation

performance, but brought significant improvements in both speech recognition and synthesis technologies as well. In addition, by using a model that converts a "Seq2Seq" sequence into another sequence, technologies for learning long-period dependencies between prior and subsequent contexts, such as LSTM*1 were invented. Researchers around the world are now developing systems based on this model.

In October 2018, Google released a new natural language processing model called "BERT*2." This is a brute-force technique, which builds a huge neural network employing enormous numbers of CPUs and GPUs for learning. This is expected to bring a giant leap in the accuracy of semantic extraction and translation.

As you may see, machine translation requires huge computing resources, and the AI world today is governed by the amount of such resources. It is safe to say that whoever has the strongest computer wins.

Social implementation of multilingual speech translation technology

——What are the social aspects of the technologies developed by ASTREC?

KIDAWARA In machine translation using neural networks, the larger the amount of data, the higher the translation accuracy. This has led ASTREC to launch the "Translation Bank" project. Global companies for instance, already possess a certain amount of their business documents translated into English and other languages. This project encourages companies to donate such documents to build up the translation data. In exchange for doing so, the license fee for using automatic translation technology is reduced considering the contents of the translation data. The Translation Bank project has seen success for pharmaceutical companies. Pharmaceutical companies must submit documents in English to the Ministry of Health, Labour and Welfare (MHLW) several times a year. It takes a month to translate the original documents into English. We fed past parallel translation data provided by a company to a neural network and found that high-accuracy translations were being generated. Although some manual corrections were necessary at the last stage, the translation took half the time, just two weeks, which was very much appreciated by the client.

Other applications of our technology include a counter service system for foreign customers developed by Toppan Printing, which is deployed in post offices all over Japan, and "Kyu-kyu (ambulance service) VoiceTra," a multilingual speech translation app that supports communication between the emergency service personnel and foreigners that have fallen ill or injured.

Future prospects

——What are the next targets of your research?

KIDAWARA We hope to realize simultaneous language interpretation by 2025 when the Expo 2025 Osaka, Kansai will be held. A prototype system is already in use, but the technology needs further research, including considering prior and subsequent contexts and estimating subjects and pronouns.

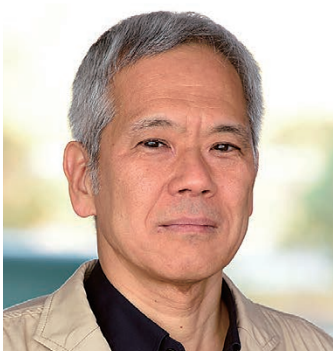
By 2030, another five years ahead, we are planning to build a sophisticated translation system that can be used for tough negotiations by diplomats and businesspeople. Developing such technologies requires skilled human resources. We are achieving great results with very few staff compared to giants like Google who can commit massive human resources. By capitalizing on our position as a national research institute, we are designing a scheme to recruit excellent students. Our laboratories have enormous computing resources and data that cannot be found in any university labs. If more students are attracted by these facilities, then we can acquire outstanding human resources.

Eliminating language barriers has been a dream of humankind for hundreds of years, and it is finally becoming a reality. There are no other fields of research that are as intriguing as multilingual speech translation. We hope that many young people will take an interest in the R&D of speech translation technologies at NICT.

*1 LSTM: Long Short Term Memory

*2 BERT: Bidirectional Encoder Representations from Transformers

Development of Multilingual Speech Recognition Technologies —Executing the Global Communication Plan



KAWAI Hisashi

Director of Advanced Speech Technology Laboratory

Advanced Speech Translation Research and Development Center

He joined the Kokusai Denshin Denwa Co. Ltd. (currently KDDI Corporation) in April 1989, the same year after obtaining his doctor's degree. He has been engaged in research on speech synthesis and speech recognition at KDDI Research, Inc. From 2000 to 2004, he was seconded to the Advanced Telecommunications Research Institute International (ATR) and engaged in research on speech synthesis. He was transferred to NICT in October 2014 and has been in his current position since. Ph.D. (Engineering).

The Advanced Speech Technology Laboratory conducts research and development of practical 1) speech recognition and 2) speech synthesis technologies for 10 languages* as part of the social implementation plan targeted for the 2020 Tokyo Olympic and Paralympic Games, 3) real-world speech recognition technologies, and 4) multilingual speech dialogue technologies targeted for the years beyond. Due to space limitations, this section will focus on the R&D outcomes of 1.

Overall structure of speech recognition systems

A speech recognition system consists of a speech recognition engine, an acoustic model, a language model, a pronunciation dictionary and external dictionary, as shown in Figure 1. While the recognition engine can be shared with any target language, the other elements must be developed for each language.

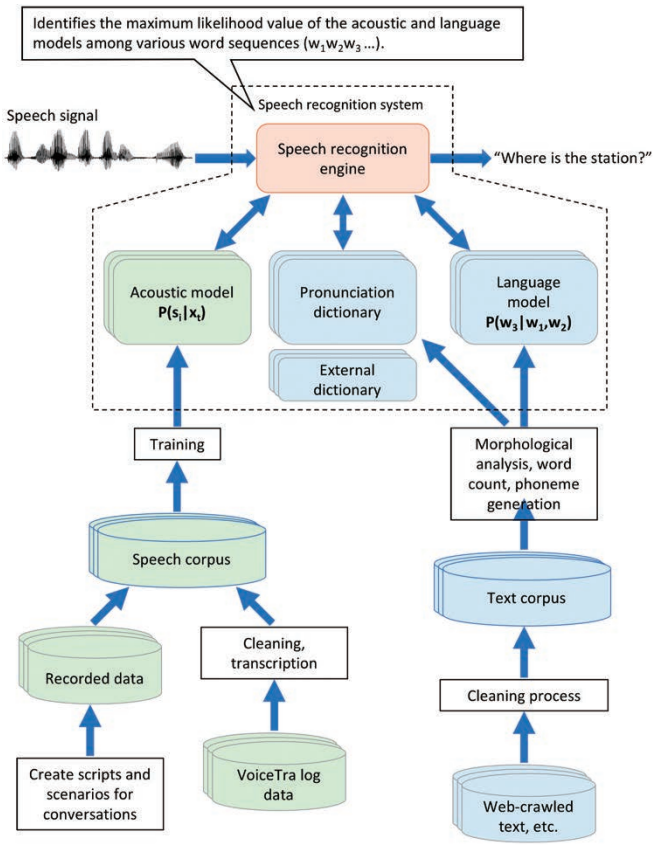


Figure 1 Speech recognition system and corpus

DNN acoustic model

An acoustic model is a model that calculates the probability $P(s_i|x_t)$ of segment x_t—about 0.01 seconds of speech signal—being the phoneme s_i. NICT's acoustic model employs a Deep Neural Network (DNN), which is created by deep learning from a large-scale speech corpus. Phonemes are the minimum units of speech, and Japanese has about a total of 40 phonemes consisting of vowels and consonants. A corpus is a large-scale collection of compiled data such as speech or text.

The architecture of the acoustic model was quite simple as shown in Figure 2 (1) around the time when the Ministry of Internal Affairs and Communications announced the Global Communication Plan (GCP) in 2014—only taking into account the speech input from the current time to output the results. As a result of continuous improvement, the structure now looks like Figure 2 (2) and uses time information from before and after the input. The use of such information is highly effective since speech itself has a temporal structure. With the improvement of the acoustic model, the word error rate which represents the speech recognition accuracy has also been reduced by as much as half.

Speech corpus

A speech corpus is a set of speech and its manual transcription which

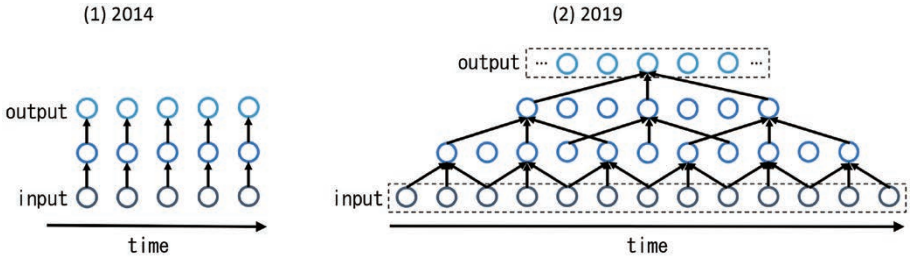


Figure 2 Advancement of DNN-based acoustic models

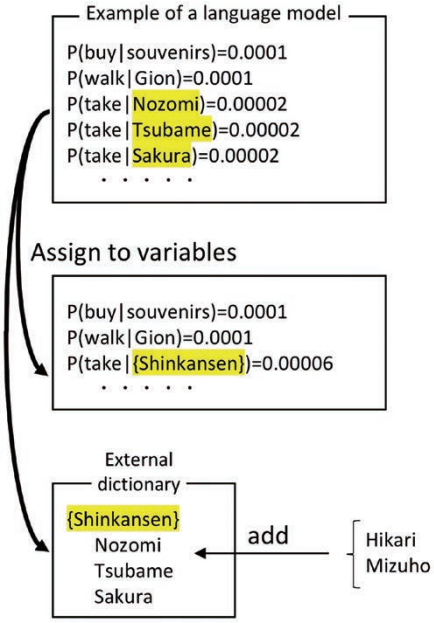


Figure 3 Example of n-gram language model and parametrization

has been carefully labeled to distinguish pronunciation errors, fillers (interjections, etc.), and pronunciations of heteronyms. A speech corpus of about 1,000 to 2,000 hours of speech is prepared for each target language. Log data is also used for languages which are frequently used in VoiceTra such as Japanese, English, and Myanmar, but for other languages recorded audio from assigned speakers are used; i.e., recordings done in conference rooms or outdoors, etc. using scripted scenarios. Training acoustic models using log data from VoiceTra showed better results when applying speech recognition to apps designed for short conversations that are 5 to 10 words long. For languages with fewer log data, it is necessary 1) to add the logs to the training data at a certain amount, 2) make improvements to the speech recognition accuracy and collect more users, 3) increase the amount of log data, 4) repeat from 1, and create a virtuous circle.

When using log data, we first eliminate the utterances of non-native speakers manually. In recorded data, the number of speakers amount to about 6 per 1 hour of recording. Since transcription errors have an adverse effect on the speech recognition accuracy, the quality of speech corpus is strictly controlled by native speakers of each target language and the number of defective utterances is kept to less than 10%.

Language model

Given a sequence of 2 words w_1, w_2 in an utterance, the language model provides the probability $P(w_3|w_1, w_2)$ of the following word being w_3. Figure 3 shows an example of a language model. A language model is created by counting a chain of 1-3 words that appear in a text corpus. For a sequence of 3 words that do not appear in the text corpus, the probability $P(w_3|w_1, w_2)$ is estimated by combining the occurrence probability of a 2-word sequence and that of the isolated word. For languages where word boundaries are not specified such as Japanese, it is necessary to perform a process called morphological analysis to estimate them.

The amount of transcription data from the speech corpus is insufficient for a text corpus, thus text data from the web—automatically collected using a web crawler—is also used. It is important to collect however, a large amount of texts that contain expressions from the field to which speech recognition is going to be applied. In addition to the cleaning process such as removing HTML tags, etc., identifying the language is indispensable when using a web-crawled corpus.

Pronunciation dictionary and external dictionary

In order to enable speech recognition for a new word, the language model must be rebuilt to add a large amount of sentences that contain the new word to the text corpus, since words that are unknown to the model cannot be recognized. However, the amount of proper nouns is infinite and new words are created every day, therefore the process of adding new words must be simple when it comes to practical use. This can be done by making the words contained in the language model into a variable. Specifically, the 3 words marked in yellow in Figure 3 are grouped into a variable {Shinkansen}, and the members of the variable "Nozomi," "Tsubame," and "Sakura" are registered in the external dictionary. The probability value of a member is basically equal, but it can also be weighted. Adding a new shinkansen name can be done by simply adding it to the external dictionary, and the language model does not need to be updated.

In addition to orthographic information, the words in the dictionary for language models are also presented using phonetic symbols. This allows the language model to be matched with the input speech via the acoustic model. Although it is ideal to specify accurate pronunciations by hand, an automatic generation process called G2P (Grapheme-to-Phoneme) is applied when dealing with large amount of vocabulary. If there are multiple pronunciations, the one that best matches the input speech is automatically selected.

Implementation to society

Our lab has already developed speech recognition systems for the mentioned 10 languages which are suitable for use in environments where the SNR (signal-to-noise ratio) is 5 dB or more. They have been licensed to private companies and are being used in various speech translation services and apps. Furthermore, we are currently developing speech recognition systems for Brazilian Portuguese, Filipino, and Nepali languages which are of high demand in Japan.

＊ 10 languages: Japanese, English, Chinese, Korean, Thai, Indonesian, Vietnamese, Myanmar, Spanish, and French

Development of Multilingual Translation Technologies —Executing the Global Communication Plan

SUMITA Eiichiro

Director of Advanced Translation Technology Laboratory

Advanced Speech Translation Research and Development Promotion Center

After receiving his master's degree from graduate school, he worked for IBM Japan and Advanced Telecommunications Research Institute International (ATR) before joining NICT in 2007. He has been engaged in research on multilingual speech translation and e-Learning. Ph. D. (Engineering).

The Advanced Translation Technology Laboratory promotes 1) research and development of practical automatic translation technologies for 10 languages as part of the social implementation plan targeted for the 2020 Tokyo Olympic and Paralympic Games, as well as the advancement of such technologies targeted for the year 2021 and beyond, 2) research on minimizing the dependence on bilingual data, and 3) fundamental research on simultaneous interpretation. This part of the article focuses on 1 and 3; topics which are rather straightforward.

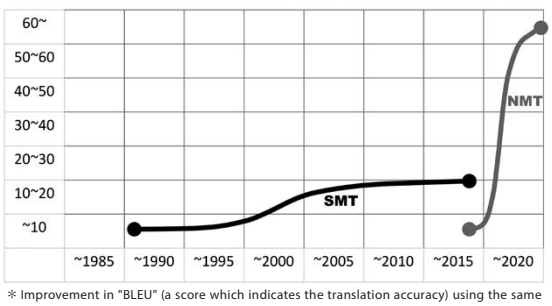
The Mechanism of automatic translation

The latest method of automatic translation uses a large-scale bilingual data (a group of input sentences in the source language paired with the output translation sentences in the target language) to learn translation knowledge. In recent years, deep-learning techniques—the basic technology behind AI—have enabled neural networks to represent translation knowledge (this technology is called "neural machine translation (NMT)"). The method of learning from bilingual data was developed in the 1980s and ever since, a number of researchers around the world have come up with various algorithms and made improvements in translation accuracy (this technology is called "statistical machine translation (SMT)"). Lately however, the rate of improvement in SMT has slowed down while NMT surpassed its performance level overnight and is making continuous improvements even as we speak (Figure 1).

A public website "Min'na no Jido Hon'yaku@TexTra"

The automatic translation system developed at NICT has been released to the public as a free field experiment app called " VoiceTra*1 " and is being widely used not only by individual users but by private companies in a licensed form for their own apps and dedicated devices. VoiceTra also plays big role in our social implementation plan targeted for the 2020 Tokyo Olympic and Paralympic Games as the app has reached a practical level of translation between 10 languages.

"Min'na no Jido Hon'yaku@TexTra*2 " (Figure 2) is a website which has also been launched as part of our field experiment and is available for free for non-commercial use. NICT's licensees on the other hand, provide



* Improvement in "BLEU" (a score which indicates the translation accuracy) using the same bilingual data (the higher the score, the better)

Figure 1 Dramatic improvement in precision achieved by new learning method—From SMT to NMT—

Figure 2 NICT's NMT is open to anyone.
https://mt-auto-minhon-mlt.ucri.jgn-x.jp/

commercial services by adding extra features, etc. to this website. "Min'na no Jido Hon'yaku@TexTra" has been localized in 4 languages: Japanese, English, Chinese, and Korean and can be used by anyone just like any other website. Users can create an account and evaluate the level of our NMT.

In addition, users may use computer-assisted translation editors without any installation, as well as Add-ins for Microsoft Office software such as Word and Excel. Also, the WebAPI can be accessed from your own script and the automatic translation engine can be launched within various computer assisted translation tools such as Trados. In such ways, "Min'na no Jido Hon'yaku@TexTra" provides a convenient set of tools concerning automatic translation.

A public website "Translation Bank"

Along with NMT, the "Translation Bank*3 " scheme is essential when it comes to practical application. Translation Bank is an approach developed by the Ministry of Internal Affairs and Communications and NICT*4 to diversify the fields in which automatic translation can be used and to increase its precision by accumulating translation data from various sources. As shown in Figure 3 for instance, we collect bilingual data related to finance, medicine, and manufacturing, and develop respective NMT systems for each field to increase the level of precision.

NMT is an excellent algorithm for automatic translation however, the translation engine cannot be developed without bilingual data; i.e., learning (training) data. NICT has been making efforts in collecting these data in cooperation with the Japan Patent Office, etc. and Translation Bank is highly effective in accelerating such approaches.

Broadly speaking, training of an NMT can be done by translating the source sentence once with the current system and by optimizing the parameters of its neural network based on the degree of difference between the translation result and the learning data. By repeating this process with huge amounts of bilingual data, an all-purpose NICT system—which serves as the common foundation for various systems—can be created.

The next step is adaptation, where additional configuration of parameters is carried out on the all-purpose system using the bilingual data for the target domain. By implementing additional parameter optimization on top of the existing all-purpose NMT, developing high-precision, tailor-made systems for different fields can be realized with relatively small amounts of bilingual data from each application domain. Table 1 shows an example of adaptation for medicine; you may see the overwhelming difference in quality.

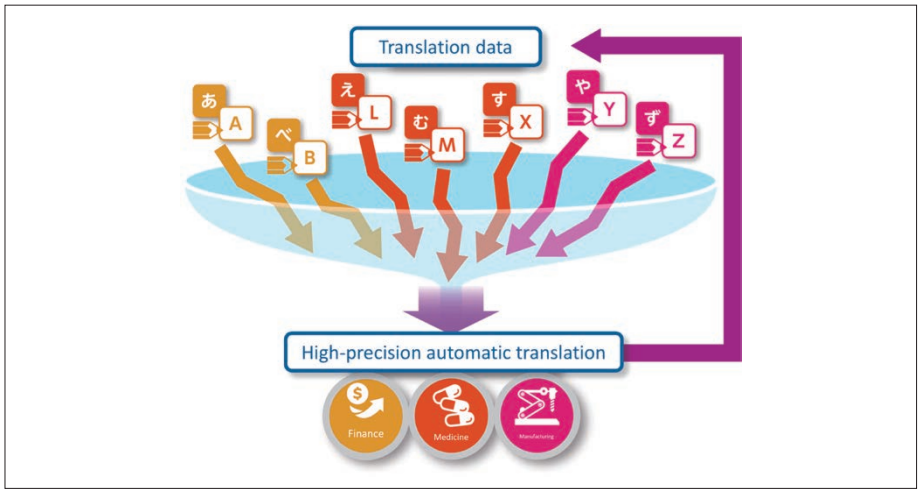


Figure 3 Outline of Translation Bank

Table 1 Example of adaptation (medicine)

Source sentence	A majority of subjects had AEs that were mild or moderate in severity and not related to investigational product (IP).
All-purpose NMT	大多数の被験者は、重症度が軽度または中等度のAESを有し、治験薬(IP)とは無関係であった。
Adapted NMT	大多数の被験者では、軽度又は中等度のAEが認められ、治験薬との因果関係は否定された。

Incrementalization of translation toward simultaneous interpretation

NICT has initiated research and development of an automatic simultaneous interpretation system using deep-learning and a prototype has already been developed. As one speaks, for instance in English, the Japanese translation is output as soon as it reaches a point where translation is available with the group of words spoken. However, for example, in order to determine whether translation can be processed for a sentence starting out with "I'd like a cup of tea" the system would need to wait until the following words are revealed; if it's "and cake" the translation must include these words, but if nothing follows, the translation can be completed. In such way, the system needs to consider information such as context, grammar, or pauses to determine the segment of words available for translation.

As delays will be caused if the translation is processed after each sentence is finished, simultaneous interpretation would become effective—as the name suggests—if the translation results can be output while one is speaking. While we will continue to make improvements to the system, it is way easier for a lot of people to recognize words in their own language even if mistranslation is involved, rather than having to comprehend a foreign language. By using simultaneous interpretation technology, TV conferences can be held between different parts of the world and both parties can speak in their native languages. The technology will bring more opportunities to work with people abroad.

Conclusion

NICT will continue to promote research and development of state-of-the-art NMT technologies and its implementation in society by utilizing bilingual data collected through Translation Bank. We kindly ask for your cooperation.

*1 https://voicetra.nict.go.jp/
*2 https://mt-auto-minhon-mlt.ucri.jgn-x.jp/
*3 https://h-bank.nict.go.jp/
*4 Jointly operated by Advanced Speech Translation Research and Development Promotion Center (ASTREC) and AI Science Research and Development Promotion Center (AIS) within NICT.

Software Development and Service Operation —Social Implementation of Research Results

ASHIKARI Yutaka

Director of System Development Office

Advanced Speech Translation Research and Development Promotion Center

After completing the master's program in graduate school, he worked for several software development companies before joining the Advanced Telecommunications Research Institute International, Inc. (ATR) in 2001 and NICT in 2006. He has been engaged in research and development of speech translation systems.

The System Development Office (SDO) carries out development of speech translation systems by integrating the research results from Advanced Speech Translation Research and Development Promotion Center (ASTREC). As part of the social implementation process, we have released our speech translation system to the public as a smartphone app called "VoiceTra." We not only keep the app up-to-date and commercial-ready, we are also committed to developing next-generation systems and apps to encourage practical use of the technologies. Exhibitions of these future systems are held in various occasions.

Multilingual speech translation app "VoiceTra"

VoiceTra (Figure 1) is a multilingual speech translation app that translates spoken words and reads them out. For example, the user can speak in Japanese and the translation in English will be read out in a synthesized voice. VoiceTra was publicly released in July 2010, becoming the world's first cloud-based speech translation app. The service operation of the app was later transferred to a private company, but in May 2015, NICT took back control and is in charge since. VoiceTra translates 31 languages, mainly Asian, of which 18 languages support speech input and 16 support speech output.* In October 2019, the language identification feature was added, which automatically determines in which language the words are being spoken. This allows a user to trigger conversations without knowing which language the other person would speak. As of December 2019, this feature can identify eight languages and more to be supported in the near future.

While serving as a demonstration app to showcase our research results, VoiceTra has another important role; to collect data. The app has marked a total of 5.6 million downloads as of November 30, 2019, and more than 200 million utterances (number of speech input) have been collected. This large volume of data is being utilized for AI learning to improve the system's performance.

Development and operation of a speech translation server system

For a cloud-based speech translation system to operate, the speech recognition, speech synthesis, machine translation, and language identification engines developed by the laboratories of ASTREC need to be implemented on a server. It is also necessary to facilitate a mechanism which enables communication between the server and the app via the Internet to process requests such as "translate Japanese into Korean and read it out in a female voice" using the appropriate

Figure 1 Screenshot of VoiceTra

* https://voicetra.nict.go.jp/

engines and to return the results back to the app. The speech translation server software is capable of handling this.

SDO is not only committed to developing high-quality server software that are ready to be commercialized, but to promoting social implementation of the research results by conducting quality testing of the engines developed by each laboratory to ensure the quality of the entire speech translation system.

Promotion of application systems

VoiceTra is one of many application examples of using the speech translation technology. SDO has been developing a variety of demonstration systems and exhibiting them in many occasions to stimulate the interest of service providers and system developers. The following are some of the systems that we have developed.

(1) Earphone-type speech translation system

This system enables communication between two people who speak different languages by using earphone-type headsets which would recognize one's language and then play the translation in the other person's language through their earphones. It appears as if for example, a Japanese person speaking in Japanese and an American person speaking in English are holding a conversation in their own languages (Figure 2). Currently, the headset is connected to each person's smartphone which handles the translation, however, development of an independent earphone-type translation device is anticipated in the near future.

(2) Real-time subtitling system

Several types of subtitling systems that display either the speech recognition results or translated text as subtitles are under development. Figure 3 shows an example of the system being used for a Power Point presentation slide—as the presenter speaks in Japanese, the English subtitles are generated and displayed at the bottom in real time. This system could be handy in international conferences with attendees who do not share the same language. The speech recognition part can be used alone without translation and Japanese for instance, has already reached a practical level which can be used for systems that provide information support for the hearing impaired. We are planning to collaborate with schools for the deaf to conduct field experiments in addition to demonstrations.

Future prospects

The scope of development will shift from speech translation systems that are capable of translating relatively short sentences one by one, to simultaneous interpretation of longer texts as seen in speeches. While each laboratory conducts further research and development of the engines used for this technology, SDO will focus on developing a new server software. The use of contextual and multimodal information must be considered in order to increase speech recognition and translation accuracies, and a universal framework to handle such information will also be required. The information required by the engines and their storage methods are yet to be determined, but we intend to develop an easily expandable and flexible platform that fully exploits the performance of the engines.

Figure 2 Earphone-type speech translation system

Figure 3 Real-time subtitling system

Social Implementation of Multilingual Speech Translation Technologies —Executing the Global Communication Plan

UCHIMOTO Kiyotaka

Director of Planning Office

Advanced Speech Translation Research and Development Promotion Center

After receiving his master's degree, he joined the Communications Research Laboratory, Ministry of Posts and Telecommunications (currently NICT) in 1996. He has been engaged in research and development and field experiments on natural language processing and speech translation, and has been committed to implementing the outcomes to society. He has been in his current position since 2015. Ph.D. (Informatics).

Recently, the number of inbound tourists to Japan is increasing and is estimated to exceed 40 million in 2020. Under such circumstances, the Ministry of Internal Affairs and Communications (MIC) announced the Global Communication Plan (GCP)*1 in April 2014, which aims to facilitate global exchanges by eliminating the language barriers of the world. To achieve this goal, NICT, together with private companies is working to improve our multilingual speech translation technology and increase the number of languages and fields supported and has been conducting field experiments and implementing the outcomes to society. Using such technology, our ultimate goal is to free the world from language barriers.

Mechanism of speech translation

NICT's multilingual speech translation technology has been released to the public as a network-based speech translation app "VoiceTra®*2," which is available on the App Store and Google Play. Supporting translation between 31 languages, VoiceTra is most useful for travel-related conversations. As Figure 1 shows, speech input via the smartphone is sent to the server through the network for speech recognition, machine translation, and speech synthesis processes, then the translated speech is returned to the smartphone via the network and played back in synthesized audio. Each process done on the server uses a mechanism that statistically learns from a language database (corpus). The precision of speech translation heavily depends on the quantity and quality of the corpus. NICT has achieved high precision by narrowing the target fields and efficiently building a corpus with the help of the VoiceTra usage log. With the target year of 2020, we have been expanding the supported areas from travel to daily life, disaster response, and the medical field.

Industry-academia-government partnership

To contribute to the promotion of the Global Communication Plan, the Council for Global Communication Development and Promotion*3 was established in December 2014 and the Multilingual Speech Translation Consortium for Commissioned Research and Development by the Ministry of Internal Affairs and Communications*4 was

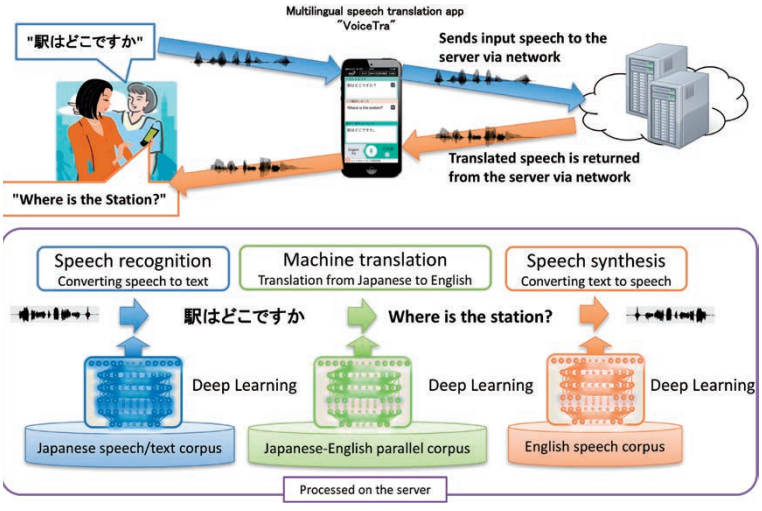


Figure 1 Outline of network-based multilingual speech translation

(Following pages are all in Japanese except *2)

*1 https://www.soumu.go.jp/main_content/000285578.pdf
*2 https://voicetra.nict.go.jp/en/
*3 https://gcp.nict.go.jp/
*4 https://www.nict.go.jp/press/2015/10/26-1.html
*5 https://news.kddi.com/kddi/corporate/newsrelease/2019/11/12/4134.html
*6 https://pr.fujitsu.com/jp/news/2017/09/19.html
*7 https://www.hitachi-solutions-tech.co.jp/iot/solution/voice/Ruby_Concierge/railway_index.html
*8 https://www.keikyu.co.jp/company/news/2017/20180328HP_17271TS.html
*9 http://www.madoguchi-honyaku.jp/
*10 https://gcp.nict.go.jp/news/products_and_services_GCP.pdf
*11 https://www.toppan.co.jp/news/2018/05/newsrelease1805141.html
*12 https://iamili.com/ja/
*13 https://panasonic.biz/cns/invc/taimenhonyaku/
*14 https://pocketalk.jp/
*15 https://www.konicaminolta.jp/melon/
*16 https://www.toppan.co.jp/news/2018/04/newsrelease1804252.html
*17 https://miraitranslate.com/uploads/2019/04/MiraiTranslate_MultilingualPlatform_pressrelease_20190426.pdf
*18 https://fairydevices.jp/price
*19 https://jpn.nec.com/cloud/service/platform_service/multilingual/index.html
*20 https://h-bank.nict.go.jp/

established in October 2015. Through these schemes, NICT has collaborated with private companies to conduct research and development, and field experiments aiming for social implementation in terms of both hardware and software, and has performed technical verifications of various user interfaces (UIs) in different scenes. For taxi applications, field experiments were conducted in Tottori City, Tokyo Metropolis, and Naha City using a speech translator developed by KDDI that can be used by both the driver in the front seat and the passengers in the rear seats*5. In the medical field, clinical trials and field experiments were conducted in hospitals and nursing homes using a wearable credit card-sized handsfree speech translation device*6 developed by Fujitsu Laboratories. In railway transportation, Hitachi Solutions Technology developed and commercialized a new multifunctional translation app, "Eki (station) Concierge®*7"—which uses the results from the collaborative research by Keikyu, Bricks, and the Hitachi Group*8—and its full-scale introduction to all Keikyu stations (except for Sengakuji Station) was completed in July 2018. The app was specifically designed for its use by supporting frequently-used terms, creating a dedicated UI, and combining with telephone interpretation services.

A joint research with the National Research Institute of Fire and Disaster of the Fire and Disaster Management Agency (FDMA) produced "Kyu-kyu (Ambulance Service) VoiceTra," an app developed based on VoiceTra with an additional feature that allows the use of pre-registered fixed phrases. As of October 1, 2019, the app has been implemented in 476 out of the total 726 fire departments (65.6%) in all 47 prefectures of Japan. Furthermore, the police departments in 29 prefectures have started test trials using VoiceTra. Led by the Okayama Prefectural Police Department, more departments are starting to develop their original apps with on-premises servers.

The demand in local governments to provide language support for foreign nationals in Japan are also on the rise. Under the commissioned research entitled "Research and Development of the Speech Translation System for Local Government*9," we have created a corpus and speech translation server dedicated for counter services at local government offices, and developed a prototype app which was used in field experiments conducted at Maebashi City, Itabashi Ward, and Ayase City Offices.

Social implementation of multilingual speech translation technologies

A number of dedicated products and services*10 for various fields and scenes have been commercialized (Figure 2). To name a few, "VoiceBiz®*11," a speech translation app from Toppan Printing; "ili*12," an offline speech translator from Logbar; a multilingual speech translation service from NEC; "Taimen Honyaku (face-to-face translation)*13" from Panasonic; "POCKETALK® W/POCKETALK® S*14," a cloud speech translator from Sourcenext; and a medical interpretation tablet, "MELON*15," from Konica Minolta. Toppan Printing's speech translation app has been introduced to approximately 20,000 post offices (except for branch offices) in Japan as "Japan Post's Yubinkyoku Madoguchi Onsei Honyaku (Post Office Counter Speech Translator)*16." Following the field experiments at their service counters, this technology is now being disseminated to municipalities. These products and services utilize the multilingual speech translation technology licensed from NICT. In April 2019, Mirai Translate started providing a multilingual speech translation platform service and licensing speech translation software*17. Speech translation API services are also provided by Fairy Devices*18 and NEC*19.

Future development

Further improvements in performance and usability are required for multilingual speech translation technology to be ubiquitous. The keys to higher precision are the collection and construction of a large-scale, high-quality corpus. NICT has started operating the "Translation Bank®*20" in which a parallel corpus is being collected under industry-academia-government collaboration with the intention of reducing costs, increasing supported fields, and accelerating technical precision. Concerning the applications of multilingual speech translation technology, it is important to combine it with different technologies such as maps, or seamlessly connect with telephone interpretation services depending on the target situation.

Amid rapid globalization, the needs for speech translation technology are expected to further grow in various scenes including business discussions and lectures at international meetings, and the MIC is currently in the process of developing the next Global Communication Plan to facilitate such needs. Geared toward the Expo 2025 Osaka, Kansai, NICT is working to realize a practical high-precision and short-latency simultaneous interpretation system by gathering various information from contexts, and is planning to implement the outcomes to society.

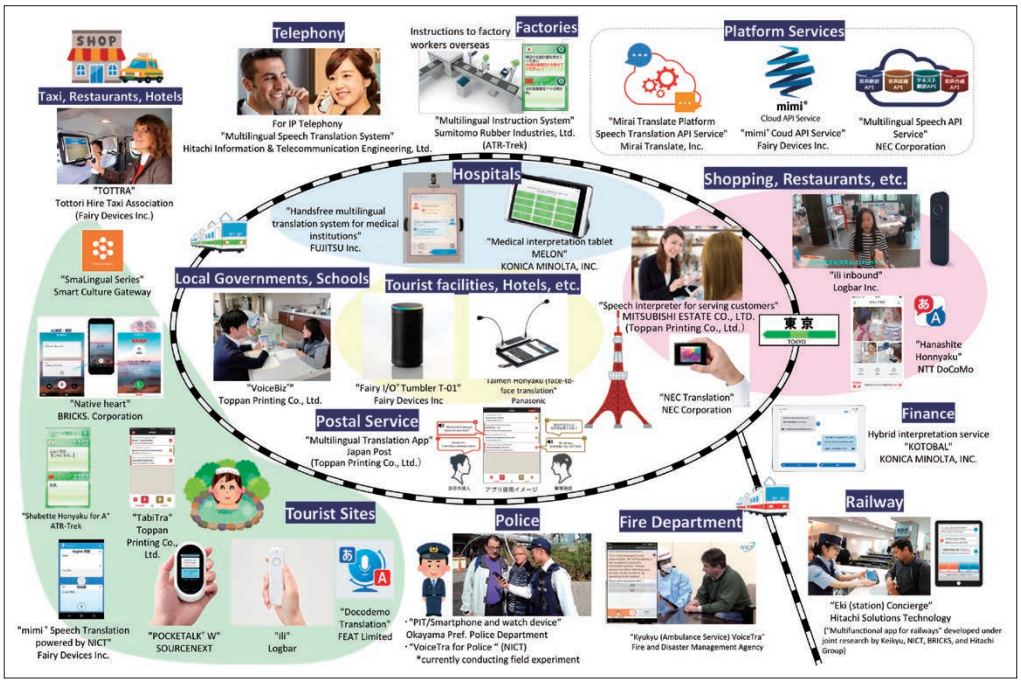


Figure 2 Social implementation examples of NICT's multilingual speech translation technology

265 nm Deep-Ultraviolet LED with an Extremely High Optical Output of over 500 mW

INOUE Shin-ichiro
Advanced ICT Research Institute,
Director of DUV ICT Device Advanced Development Center

Light is a type of electromagnetic wave with a broad range of wavelengths. Deep-ultraviolet (DUV) light with wavelengths of 200 to 300 nm is classified as light with the shortest wavelengths that can propagate through air. Deep-ultraviolet light with a wavelength of less than 280 nm (UV-C region), which is not present in natural sunlight (on the surface of the Earth) because it is completely absorbed by the Earth's ozone layer, has special properties. For example, viruses and bacteria that have evolved in an environment without exposure to DUV light have a structure that strongly absorbs light in the DUV region. Therefore, they can be sterilized very effectively and cleanly by DUV irradiation, without using chlorine or other chemicals. Since DUV light does not exist in nature, it is expected to be used in applications including telecommunications and sensing that are not affected by background sunlight noise, as well as non-line-of-sight (NLOS) optical communication using a high scattering coefficient in the atmosphere. Being the shortest-wavelength light that propagates through air, DUV light is expected to play invaluable roles in diverse technical fields such as lithographic microfabrication, 3D printing, curing for resins, photodegradation of environmental pollutants, spectral analysis, and medicine.

In industries, mercury lamps have been mainly used as DUV light sources. However, these devices are bulky and contain mercury, which is harmful to humans and has a high environmental impact. Today, there is a pressing need to develop alternatives to mercury under the Minamata Convention on Mercury (which took effect on August 16, 2017), an international agreement to reduce and ultimately eliminate the production and use of mercury in products. Accordingly, research institutes around the world have been stepping up the research and development of DUV light-emitting diodes (LEDs). However, DUV-LEDs have the disadvantage of an extremely low optical output compared to mercury lamps.

The DUV ICT Device Advanced Development Center, Advanced ICT Research Institute has been working on enhancing the output of DUV-LEDs based on nanophotonic technologies. Having conducted comprehensive development including a nano-light extraction structure, which suppresses internal optical absorption and optical output saturation (efficiency droop); device and chip electrode structures; packaging; and implementation technology, we have succeeded in demonstrating the operation of a single-chip 265 nm DUV-LED with the world's highest output of more than 520 mW (continuous-wave operation at room temperature). (Figure 1)

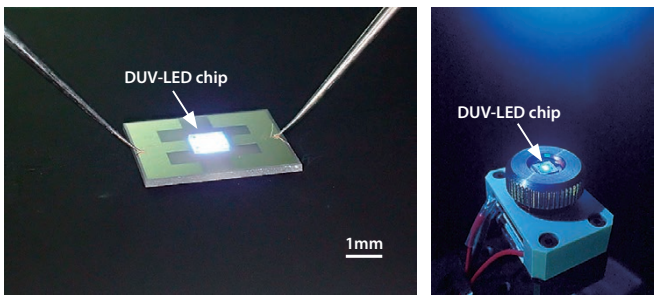


Figure 1 External view of emission of a DUVLED when a current is applied (Left: On a sub-mount; Right: Implemented)

In addition to serving as an environmentally friendly alternative to mercury lamps, our DUV-LEDs enable compact, portable, and high-output devices to be developed for handy virus sterilization systems, point-of-care medicine, and home electric appliances. Using the features of a small, low-voltage solid-state DUV light source that does not require warming up, a variety of entirely new applications are expected. To spread the use of this technology in society, we have been forging government-industry partnerships with private companies. Going forward, to rapidly disseminate the technology and contribute to the creation of new industries, we will continue researching and developing DUV optical devices including DUV light control devices, aiming for application to epoch-making optical information and communication technology (ICT) such as NLOS light communication and DUV sensing, while enhancing performance and reliability.

Our research results were highlighted on the cover of the October 2019 issue of the *Japanese Journal of Applied Physics*, published by the Japan Society of Applied Physics (Figure 2).

Figure 2 Cover of the October 2019 issue of the *Japanese Journal of Applied Physics* showing a scanning electron microscope (SEM) picture of the nanophotonic structure formed on the developed DUV-LED, with the author's comments

INOUE Shin-ichiro

After receiving a Ph.D. degree, he worked for RIKEN, and Kyushu University. Entered NICT in 2010. He is also a visiting professor of Graduate School of Engineering, Kobe University. Ph.D. (Engineering).

Multilingualization of Computational Linguistics and Natural Language Processing Centered Around Asian Languages

Biography

1986 Born in Jinan, Shandong, China
2009 Graduated from the School of Mathematics and Statistics, Shandong University, China
2012 Completed the master's program at the Graduate School of University of Tsukuba
2015 Completed the doctor's program at the Graduate School of University of Tsukuba and joined NICT
2018 Current position

Awards, etc.

PACLING 2017 Best Paper Award

DING Chenchen

Researcher (Tenure-Track)
Advanced Translation Technology Laboratory
Advanced Speech Translation Research and Development Promotion Center (ASTREC)
Ph.D. (Engineering)

Q&As

Q If you were reborn, what would you like to be?

A I would like to have several mother tongues, which perhaps may widen my view and understanding of the world. I would learn the most typical ones from each language category—isolated languages (Chinese, etc.), agglutinative languages (Japanese, etc.), and fusional languages (German, Russian, etc.)—from the viewpoint of linguistic typology.

Q What are you currently interested in outside of your research?

A I started playing the clarinet again in 2018, after a hiatus of thirteen years. I am now practicing Carl Maria von Weber's Concertino for Clarinet (Op. 26).

Q What advice would you like to pass on to people aspiring to be researchers?

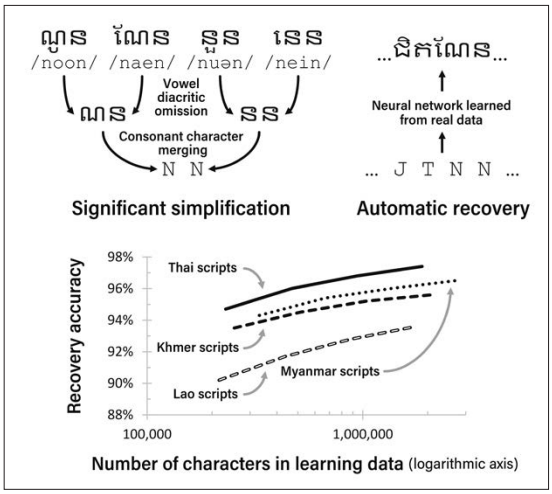
A "The accomplished scholar is not a utensil." (from the Analects, Book II.)

There are some 6,000 languages that exist in the world. However, the targets of research on natural language processing technology tend to be biased toward European languages and just a few Asian languages. Researches on the rest of the many languages are immature and untapped. Moreover, the data for these languages are usually insufficient. In order to achieve the ultimate goal of eliminating the language barriers of the world, we need to solve the mentioned issues for these "low-resource languages."

However, it is not realistic to target all of the world's languages. Considering the social impact and the economic effects, it is unquestionable that the emphasis of research at NICT—a national institute of Japan—should be placed on Asian languages. My mission is to build a language corpus mainly consisting of Asian languages and to develop a versatile language analysis technology. I would like to raise the bar for Asian language processing to the same level as English, Japanese, and Chinese by creatively applying the processing technologies used for them, which have a long development history of more than 70 years.

One of the outcomes of my research is the development of a software called "AKKHARA," which enables users to efficiently type abugida characters that are widely used in Asian languages. Inputting characters is the basis of IT and language processing, however not all countries have developed sufficient technology and are often faced with confusions and unnecessary time consumption. To solve this problem, activities to disseminate AKKHARA to PC users in Asian countries including Myanmar and Cambodia are under way.

I believe that our research will contribute to various fields of study including computational linguistics and natural language processing, promote the development of ICT in Asia, assist Japanese companies to branch out to Asia, and ultimately contribute to the prosperity of the whole Asian continent.



In order to facilitate efficient input methods for abugidas, the characters are first simplified by reducing the redundancies in the writing system. Then, using the latest neural network technologies, the characters are recovered. The upper part of the figure shows the simplification and recovery processes with Khmer scripts in Cambodia as an example, and the lower part indicates the recovery accuracies with Thai, Myanmar, Khmer, and Lao scripts. The performance is improved as the amount of real data increases.

NICT NEWS 2020 No.2 Vol.480
Published by **Public Relations Department, National Institute of Information and Communications Technology**
Issue date: March 2020 (bimonthly)

4-2-1 Nukui-Kitamachi, Koganei, Tokyo
184-8795, Japan
TEL: +81-42-327-5392 FAX: +81-42-327-7587

URL: https://www.nict.go.jp/en/
@NICT_Publicity
#NICT
ISSN 2187-4050 (Online)