

7-4 International Standardization of the Network-based Speech-to-Speech Translation Technologies and Expansion of the Standardization Technologies by the International Research Collaborations

HORI Chiori

The rapid growth of information communication technologies and transportation has resulted in accelerating the explosive increase of interactions between the people across the globe today. However, the language barriers still hinder and interfere with people's communication. As a useful means to break these barriers, the speech-to-speech translation (S2ST) system is now drawing attentions from various fields. As components of S2ST, automatic speech recognition (ASR), machine translation (MT) and text-to-speech synthesis (TTS) for covering different language have been developed independently and separately in many research institutes of the world. Connecting the various distributed servers for these components through the network makes the S2ST for more languages enable. In order to acquire the speech translation outcome through connecting servers with various input/output through the network, it is the most imperative that the communication protocols between modules of S2ST should be internationally standardized at ITU-T (International Telecommunication Union Telecommunication standardization Sector). Therefore NICT and Asian research institutes established the Asian Speech Translation Advanced Research (A-STAR) in 2006 and launched the activities of standardization of the network-based S2ST protocol. Then the activities were shifted to the Universal Speech Translation Advanced Research Consortium (U-STAR) in line with the transfer of standardization activities of the network-based S2ST protocol. In October, 2010, the protocol standardization was approved at ITU-T as the ITU-T Recommendations, F.745 and H.625. The U-STAR is now expanding its activity with 26 member institutes from 23 countries, and has been conducting one-year field experiment by connecting the members' servers which are built with the ITU-T standardized protocol.

Keywords

Automatic speech recognition (ASR), Machine translation (MT), Text-to-speech synthesis (TTS), Speech-to-speech translation (S2ST), Universal Speech Translation Advanced Research Consortium (U-STAR)

1 Introduction

The fact that the world has many different languages is one of the barriers to people's communication. The more directly people who speak different languages can communicate without language boundaries, the more mutual

understanding can be accelerated and the closer human relationships can be constructed all over the world. To achieve such communication between humans, S2ST technologies can be used. S2ST is a technology that recognizes the speech in one language, translates the recognized speech into another language, and

then synthesizes the translation into speech. The leveraging of S2ST technologies in a pragmatic manner, which has long been one of mankind's dreams, is very much expected to make contributions to the daily scenes such as tourism, social services, safety, and security by removing language barriers, and may ultimately influence language education. To construct S2ST systems consists of ASR, MT and TTS and these modules use the constructed models by learning the data such as speeches, transcriptions, pronunciation lexica, and parallel translation corpora for each language. It is very difficult for individual organizations to build S2ST systems covering all topics and languages. However, by interconnecting ASR, MT and TTS modules developed by separate organizations and distributed globally through a network, one can create S2ST systems that break the world's language barriers. NICT established the Asian collaborative research consortium A-STAR and had initiated the standardization activities for Asian network-based S2ST protocol, at the Asia-Pacific Telecommunity Standardization Program (ASTAP: <http://www.ap.t.int/APTASTAP>) since 2006. In 2009 NICT, as a member of U-STAR, started the standardization activity at International Telecommunication Union, Telecommunication Standardization Sector (ITU-T: [http://](http://www.itu.int/ITU-T/)

www.itu.int/ITU-T/). And the standardization documents of the protocol which NICT proposed as a representative of U-STAR were approved as ITU-T Recommendations, F.745 and H.625 [1][2] in October, 2010.

Currently 26 institutes from 23 countries are signatory members of U-STAR. U-STAR now constructed the speech translation network by connecting each of the institutes' servers built on the network based S2ST protocol standardized at ITU-T and is scheduled to conduct a one-year field experiment of speech translation in 2012.

2 Network-based Speech-to-Speech Translation (S2ST) technology

2.1 Architecture of speech-to-speech translation system

The process of a S2ST system is: the source speech input is converted into the speech text by ASR, the text is resulted into the target language text by MT, and the translation text is output with synthesized voice (TTS). Figure 1 shows an example of the process flow of a S2ST system. In each module of the S2ST system, the process of speech recognition, translation, and speech synthesis is advanced by using the learning models from speeches,

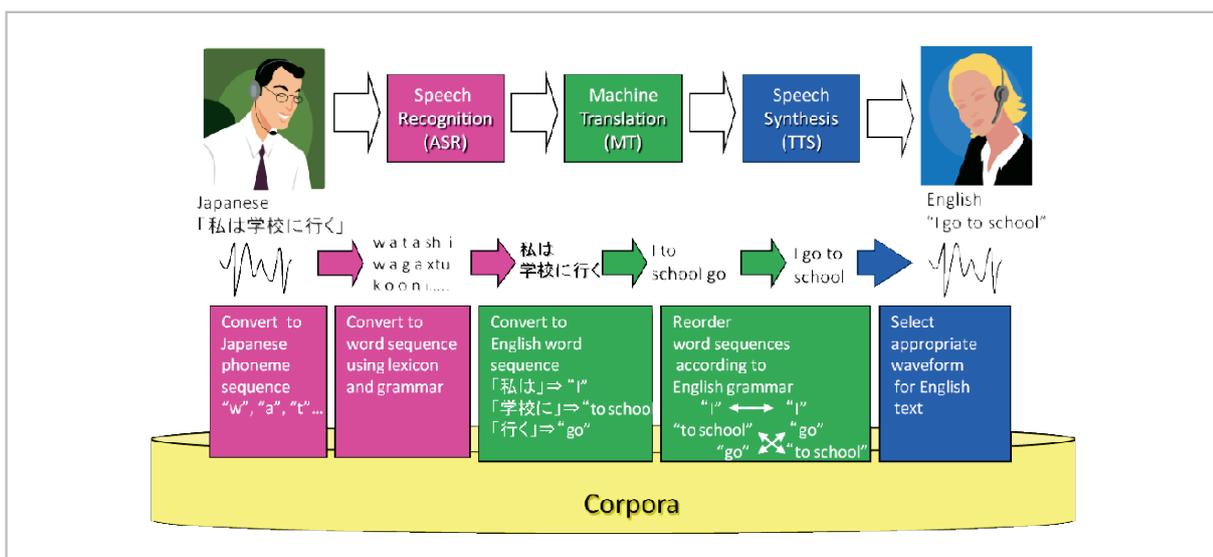


Fig.1 Architecture and process of S2ST system

transcriptions, and parallel corpora.

2.2 The network-based S2ST technologies

Through connecting the speech translation modules of ASR, MT and TTS modules distributed and developed by the research organizations all over the world through the network, it is possible for us to create the network-based S2ST systems for much more languages. In order to realize the system, we need to have the large-scale corpus such as speech data, transcriptions, pronunciation lexica, and parallel corpora for translation, all of which are required to build the acoustic model, the language model, and translation model for ASR, MT and TTS. Further importantly, the standardization of the communication protocol and data format is very much needed for connecting modules with various languages and functions reliably as illustrated in Fig. 2, which shows an example of a high level functional model of a network-based speech translation. NICT was a co-founder of the international

collaborative research consortiums with other international research institutes, and initiated the procedure of International standardization of S2ST communication protocol, collecting the large-scale learning data required for different speech and text sources for the speech translation and promoting the research activities of the speech translation technology. In the following section, the international activity and procedure to realize the standardization are described.

3 Expansion of the standardization activity — from Asia to the World —

3.1 Standardization activities start in Asian region

Research laboratories in Asia started an Asia-wide S2ST research effort by forming a consortium, called A-STAR, in November 2006. We now have materialized network-based S2ST systems in a collaboration with eight institutes from eight Asian countries i.e., NICT(Japan), ETRI(Korea), CASIA(China),

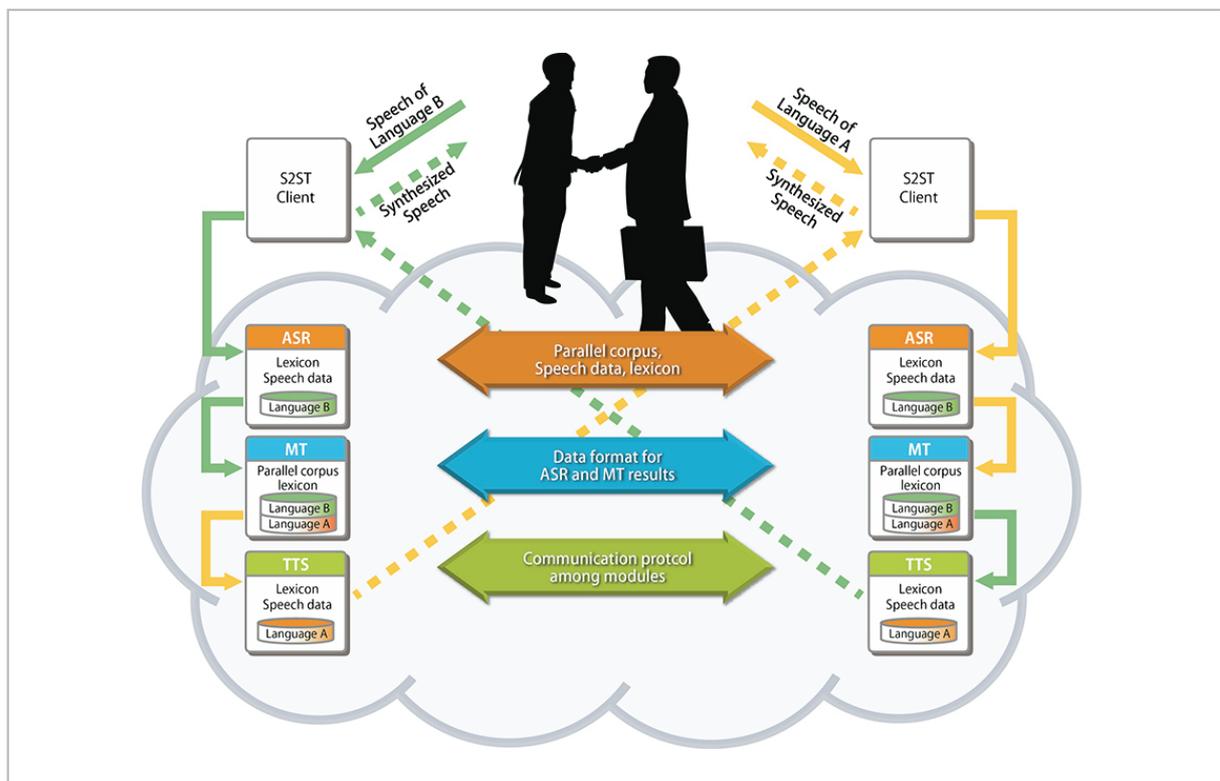


Fig.2 High-level functional model of a network-based S2ST with data flow

NECTEC(Thailand), BPPT(Indonesia), CDAC(India), IOIT(Vietnam) and I2R(Singapore), for the travel domain. And in July of 2009, A-STAR took demonstration of the network-based S2ST, connecting the speech translation modules which have been developed in each of Asian institutes through the network. This event represented that S2ST could be an effective means of communication between people who speak different languages.[3] The next and most imperative thing was to internationally standardize the communication protocol to realize network-based S2ST technologies by connecting many more various distributed modules not only in Asia but in the rest of the world. The procedure of standardization of communication protocols between modules of S2ST and data formats should be done in a prompt manner. Thus far, standardization activity of network-based S2ST was carried out in ASTAP. To extend the network of S2ST and invite more languages for translation, it is required to create a global standard for S2ST technologies and dis-

seminate S2ST technologies to the entire world. The members of ASTAP unanimously agreed at the plenary session at ASTAP15 held in March of 2009 to transfer standardization activities of S2ST to ITU-T from ASTAP [4]. Refer to the standardizing procedures how ASTAP were transferred to ITU-T as the consortium was shifted from A-STAR to U-STAR in Fig. 3.

3.2 International standardization of network-based S2ST protocols at ITU-T

In October 2009, NICT, as a member of U-STAR has initiated the standardization activity of the network-based S2ST communication protocol and the standardization procedure was taken place as SG16 (Multimedia coding, systems and application), WP2 (Applications and systems), and Q21 (Multimedia architecture)/Q22 (Multimedia applications and services). NICT, as the editor, drafted and submitted the documents of two recommendations; “Functional Requirements for the Net-

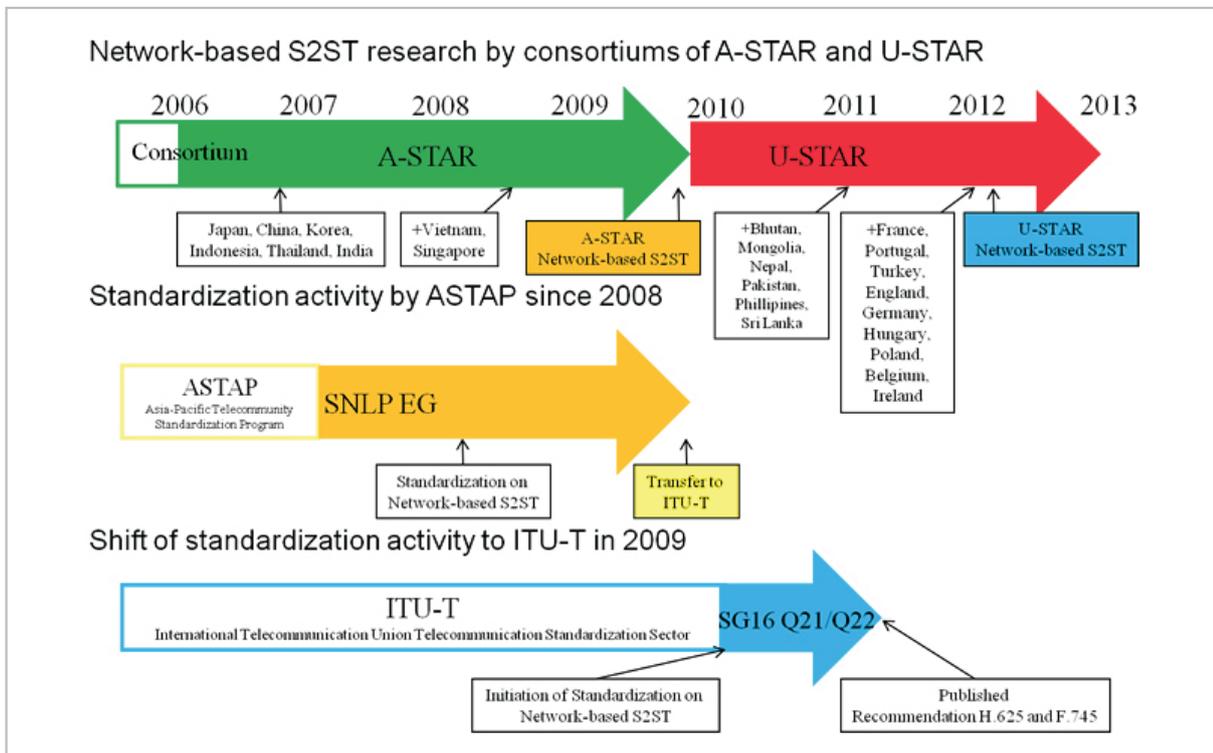


Fig.3 Expansion of the international research collaboration and standardization procedure of the S2ST communication protocol

work-based S2ST: F.S2STreqs”, and “Architecture for the Network-based S2ST: H.S2STarch”. These protocols were approved at ITU-T as ITU-T Recommendations, F.745 and H.625, in October 14th 2010, enabling S2ST modules to be connected across the globe over networks, as shown in Table 1.

3.3 U-STAR launching one-year field experiment for the London Olympics

U-STAR consortium has been expanding its activities worldwide with an increased number of members, 26 institutes from 23 countries (as of June, 2012). The language covered by the consortium has come up to 23 languages, which deserves about 95.4% of the world population. Figure 4 indicates the area map where the languages supported by

U-STAR are spoken as primary language. And Table 2 shows a list of U-STAR member institutes.

In accordance with a collaborative research by U-STAR members, NICT has been conducting the field experiment, connecting the network-based S2ST applications developed on smartphone based on the standardized protocol at ITU-T Recommendations F.745, H.625, with the speech translation servers distributed in each of U-STAR member institutes. The application developed in this effort are: a single device type of application used for conversation face-to-face at real time and a multi-device type of application and the other for a multi-device used for conversation either face-to-face or remotely. And U-STAR makes the

Table 1 Standardization of network-based S2ST protocols at ITU-T		
	Recommendation ITU-T F.745	Recommendation ITU-T H.625
Title	Functional Requirements for the Network-based S2ST	Architecture for the Network-based S2ST
Scope	It defines the requirements and architecture for connecting modules such as speech recognition, machine translation, and speech synthesis, needed for the speech translation service on the network http://www.itu.int/rec/T-REC-F.745-201010-I	It defines the functional architecture and mechanisms of network-based S2ST, interface protocols between S2ST modules, and a workflow of the network-based S2ST system. http://www.itu.int/rec/T-REC-H.625-201010-I



Fig.4 U-STAR maps of member institutes and language coverage

Table 2 List of U-STAR member institutes

 Agency for the Assessment and Application of Technology (BPPT), Indonesia	 Institute of Automation, Chinese Academy of Sciences (CASIA), China	 Center for Development of Advanced Computing (CDAC), India	 Electronics and Telecommunications Research Institute (ETRI), Korea	 Institute for Infocomm Research (I2R), Singapore
 Institute of Information Technology (IOIT), Vietnam	 National Electronics and Computer Technology Center (NECTEC), Thailand	 National Institute of Information and Communications Technology (NICT), Japan	 Department of Information Technology and Telecom (DITT), Bhutan	 Al-Khawarizmi Institute of Computer Science, UET (KICS-UET), Pakistan
 Language Technology Kendra (LTK), Nepal	 Mongolian University of Science and Technology (MUST), Mongolia	 National University of Mongolia (NUM), Mongolia	 University of Colombo School of Computing (UCSC), Sri Lanka	 University of the Philippines Diliman (UPD), Philippines
 Budapest University of Technology and Economics Dept. of Telecommunications and Media Informatics (BME-TMIT), Hungary	 National Center of Scientific Research (CNRS-LIMS), France	 Institute of Systems and Computer Engineering - Research and Development in Lisbon, (INESC-ID), Portugal	 Polish-Japanese Institute of Information Technology, (PJIT), Poland	 Pázmány Péter Catholic University, (PPKE), Hungary
 University of Sheffield, Department of Computer Science, Speech and Hearing Group, (SpandH), UK	 KU Leuven, Dept. Electrical Engineering, division PSI-Speech, (ESAT), Belgium	 Technische Universität München, (TUM), Germany	 Trinity College Dublin, (TCD), Ireland	 Center of Research for Advanced Technologies of Informatics and Information Security, (TUBITAK), Turkey
 Ulm University - Institute of Communications Engineering, (UUm) Germany				

public release of the iPhone* applications using the network-based S2ST at U-STAR Workshop in London, June 2012 [5]. U-STAR is also scheduled to launch a one-year field experiment of the network-based S2ST system for 2012 London Olympic Games.

4 Conclusion

NICT successfully constructed the framework to realize S2ST for more languages with which the various S2ST modules distributed all around the world could be connected, by leveraging the communication protocol of the S2ST standardized at ITU-T. In the growing activity of U-STAR, more and more international institutes join the consortium, advancing

the research of S2ST technologies of their own in the course of improvements and development through the field experiment via the network-based S2ST system in the pragmatic environment. In addition, transferring this speech translation technology of each of U-STAR institutes to the private enterprises in their countries may accelerate the business opportunities of the speech translation in the market. It is our hope that the network-based speech translation enables the human's long-held dream come true; overcoming the languages barriers in the world and makes a great contribution to the international society.

* 1 iPhone is a trademark for Apple Inc. and registered in the United States and other countries.

References

- 1 Recommendation ITU-T F.745 (2010), Functional Requirements for Network-based S2ST. <http://www.itu.int/rec/T-REC-F.745-201010-I>
- 2 Recommendation ITU-T H.625 (2010), Architecture for Network-based S2ST. <http://www.itu.int/rec/T-REC-H.625-201010-I>
- 3 Chairman's report of the A-STAR meeting in TCAST2009.
- 4 ASTAP09/FR15/01, "Proceedings of ASTAP15".
- 5 Public Release of the Network-based S2ST application, "VoiceTra4U-M" at the U-STAR Workshop in London June 27, 2012. <http://www.ustar-consortium.com/app/app.html>

(Accepted June 14, 2012)



HORI Chiori, Ph.D.

*Senior Researcher, Spoken Language
Communication Laboratory, Universal
Communication Research Institute
Speech Recognition, Speech Translation,
Spoken Dialog Technologies*