

7-4 ネットワーク型音声翻訳の国際標準化と国際研究協力による標準化技術普及の取り組み

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

堀 智織

HORI Chiori

要旨

近年、情報通信技術や交通の発達により、世界中の人々と交流する機会が爆発的に増えているが、“言葉の違い”が高い壁となり、人々のコミュニケーションを妨げているという問題がある。その解決策として、自分の音声を手前の言語に翻訳し、音声で相手に聞かせる「自動音声翻訳システム」が注目されている。現在、世界中の研究機関が、自国語の音声認識 (ASR)、機械翻訳 (MT)、音声合成 (TTS) などの音声処理技術の研究に盛んに取り組んでいる。これらの世界に分散配置された音声翻訳サーバをネットワークで接続することにより、より多くの言語を音声翻訳することが可能となる。但し、このような入出力の異なるサーバをネットワークで接続して音声翻訳結果を得るためには、各モジュール間の通信プロトコルを国際標準化する必要がある。NICT はアジアの研究機関と共同で「アジア音声翻訳先端研究コンソーシアム (A-STAR)」を 2006 年に設立し、ネットワーク型音声翻訳の通信プロトコルの標準化活動を開始した。A-STAR とその音声翻訳のための通信プロトコルの標準化活動は、2009 年に国際研究コンソーシアムとして新たに設立された「ユニバーサル音声翻訳先端研究コンソーシアム (U-STAR)」に引き継がれている。U-STAR は ITU-T にてネットワーク型音声翻訳の通信プロトコルの国際標準化活動を行い、提案プロトコルは 2010 年 10 月に勧告書 F.745、および H.625 として正式に国際標準技術として承認された。現在 U-STAR には 23 ヶ国から 26 研究機関が参加しており、ITU-T 標準化プロトコルを用いて世界各国のサーバを接続し、2012 年に約 1 年間に亘る音声翻訳実証実験を行っている。

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, 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 speech-to-speech translation 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 Univer-

sal 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.

[キーワード]

音声認識 (ASR), 機械翻訳 (MT), 音声合成 (TTS), 音声翻訳 (S2ST), ユニバーサル音声翻訳先端研究コンソーシアム (U-STAR)

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 はじめに

世界中には様々な言語があり、コミュニケーションを阻む壁となっている。言語を超えて相互理解し、より良い人間関係を構築するコミュニケーションを実現する方法として、音声翻訳 (Speech-to-Speech Translation: S2ST) 技術が注目されている。音声翻訳技術とは、自分の音声を認識し、認識された音声を相手の言語に翻訳し、音声で相手に聞かせるという技術である。この音声翻訳技術を実環境で用いることは、人類の長年の夢であり、言語の壁を越えたコミュニケーションが、観光、社会サービス、さらには言語教育などの場面において貢献する事が期待されている。音声翻訳システムは、ASR、MT、TTSの各モジュールで構成されている。各モジュールは、言語毎に音声データ、書き起こしデータ、発音辞書、対訳コーパスなどの音声および言語データを学習データとして用いたモデルを用いている。そのため、単一の組織が全ての言語であらゆる分野を網羅した音声翻訳システムを構築することは非常に困難である。一方、世界中に分散している各研究機関で開発された ASR、MT、TTS モジュールをネットワークで接続することにより、世界中の言語の壁を越える音声翻訳システムを作ることが可能となる。NICT は、2006 年から国際共同研究コンソーシアム A-STAR を設立し、アジア・太平洋電気通信標準化機関 (ASTAP: <http://www.apt.int/APTASTAP>) にてアジアにおけるネットワーク型音声翻訳のprotocolsの標準化を開始した。2009 年には、アジアに留ま

らず世界中の音声翻訳モジュールを接続するため、U-STAR の一員として国際電気通信連合・電気通信標準化部門 (ITU-T: <http://www.itu.int/ITU-T/>) にて国際標準化活動を開始した。2010 年 10 月、提案プロトコルが ITU 勧告書 F.745 および H.625 によって世界標準として認められた [1][2]。現在 U-STAR には 23 ヶ国の 26 研究機関が加盟しており、ITU-T 標準ネットワーク型音声翻訳プロトコルを用いて各研究機関のサーバを接続した音声翻訳ネットワークを構築している。U-STAR は、2012 年に約 1 年間の音声翻訳実証実験を予定している。

2 ネットワーク型音声翻訳技術

2.1 音声翻訳システム

音声翻訳システムは、まず ASR を用いて発話された音声信号を書き起こし、その書き起こしテキストは MT によって目標言語のテキストへ変換され、最後に TTS によって翻訳結果のテキストから合成音声を生成するという順に処理される。図 1 に音声翻訳システムの処理過程を示す。音声翻訳システムの各モジュールでは、音声、書き起こし、対訳文などの大規模コーパスから機械学習したモデルを用いて、適切な認識文、翻訳文を推定し、音声として合成する。

2.2 ネットワーク型音声翻訳システム

世界中の研究機関が開発している自国語の ASR、MT、TTS という音声翻訳モジュールをネットワークで接続することにより、世界規模の

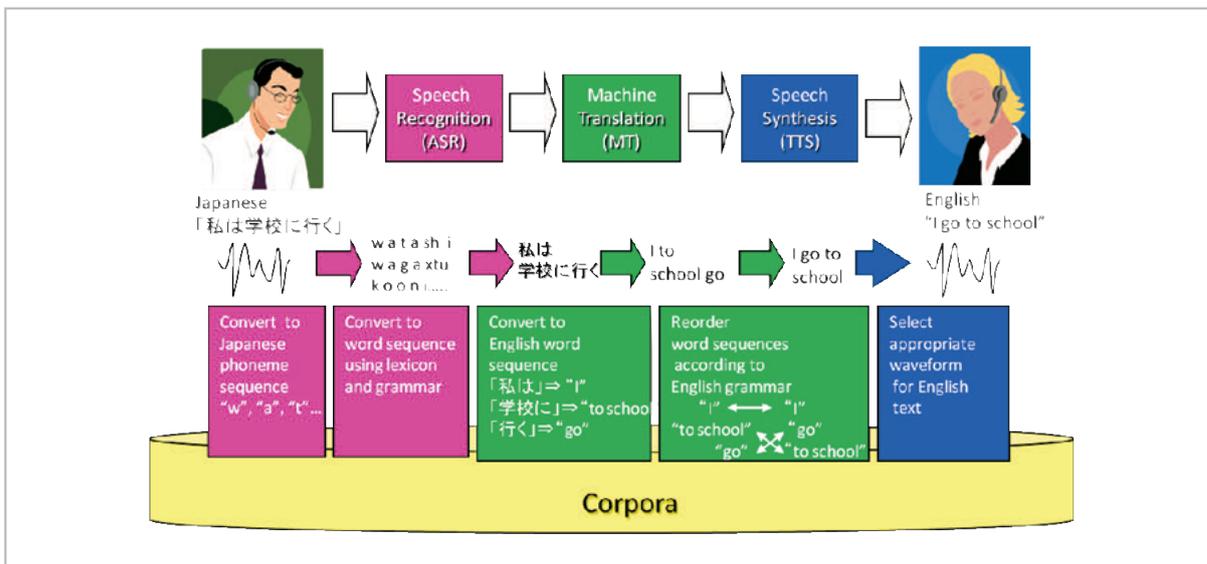


図 1 音声翻訳システム構成および処理過程

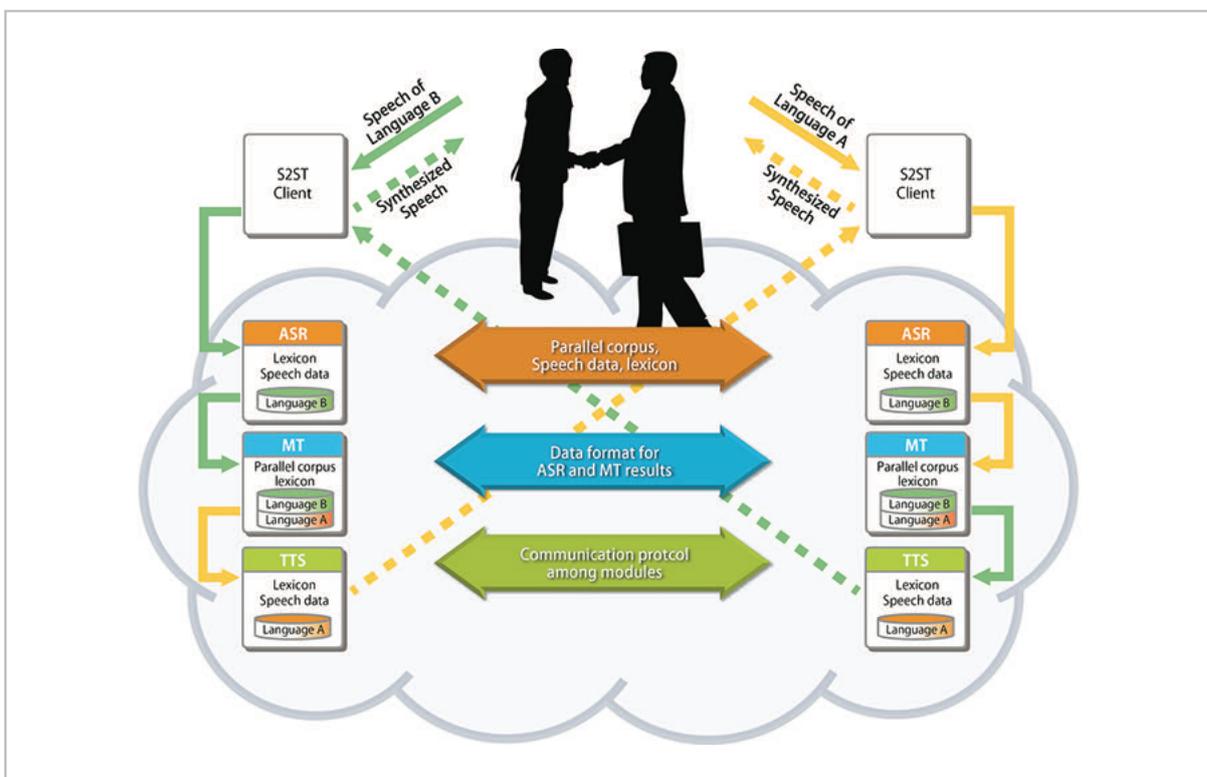


図 2 ネットワーク型音声翻訳システムの構成と処理過程

音声翻訳ネットワークを実現できるようになり、より多くの言語を音声翻訳することが可能となる。このようなネットワーク型音声翻訳を実現するためには、まず各音声翻訳のモジュールである ASR、MT、TTS で用いられる音響モデル、言語モデル、対訳モデルを構築するため、各言語の

音声データ、書き起こし、発音辞書、対訳コーパスといった音声とテキストの大規模コーパスが必要となる。さらに、入出力の異なる様々な言語のモジュールを接続するためには、図 2 に示すように、モジュール間を繋ぐ通信プロトコルやデータ形式を標準化させることが必須である。NICT

は世界の音声翻訳モジュールをネットワークで繋いで世界中の言語を音声翻訳する事を目的として、世界の研究機関と共に国際共同研究コンソーシアムを創設し、各言語の音声翻訳に必要な音声およびテキストの大規模学習データの収集を行い、音声翻訳技術の研究開発を推進し、さらに音声翻訳モジュール間の通信プロトコルの世界標準化を行った。以降の章で、国際標準化活動の詳細を述べる。

3 国際標準化活動の拡大 ～アジアから世界へ～

3.1 アジアにおける標準化活動の始動

NICT はアジアの研究機関と共同で、2006年11月アジア音声翻訳先端研究コンソーシアム(A-STAR)を設立し、アジア圏におけるネットワーク型音声翻訳の研究を開始した。NICT(日本)、ETRI(韓国)、CASIA(中国)、NECTEC(タイ)、BPPT(インドネシア)、CDAC(インド)、IOIT(ベトナム)、I2R(シンガポール)の8ヶ国8機関が協力し、主に旅行対話を対象とした

音声翻訳システムを構築した。さらに、アジアに分散配置されている各研究機関の音声翻訳モジュールをネットワーク越しに接続し、2009年7月に実証実験を行った。実験では実時間で音声翻訳対話を実現し、異言語間のコミュニケーションの手段としてネットワーク型音声翻訳システムが有効であることが証明された[3]。この音声翻訳ネットワークをアジアだけでなく世界に拡大し、より多くの音声翻訳モジュールを接続するため、モジュール間通信プロトコルの標準化が必須であり、ネットワーク型音声翻訳における通信プロトコルの標準化が急がれた。A-STARによる標準化活動は、まずアジア・太平洋電気通信標準化機構(ASTAP)において始動した。さらに、世界中にある言語の壁を越えるためには、ネットワーク型音声翻訳通信プロトコルをアジアだけでなく世界で標準化する必要がある。そこで、2009年3月のASTAP15本会議において、ネットワーク型音声翻訳技術の標準化活動をASTAPからITU-Tに移行して国際標準化していくことが全員一致で採決された。図3では、国際研究コンソーシアムがA-STARからU-STARへ移行

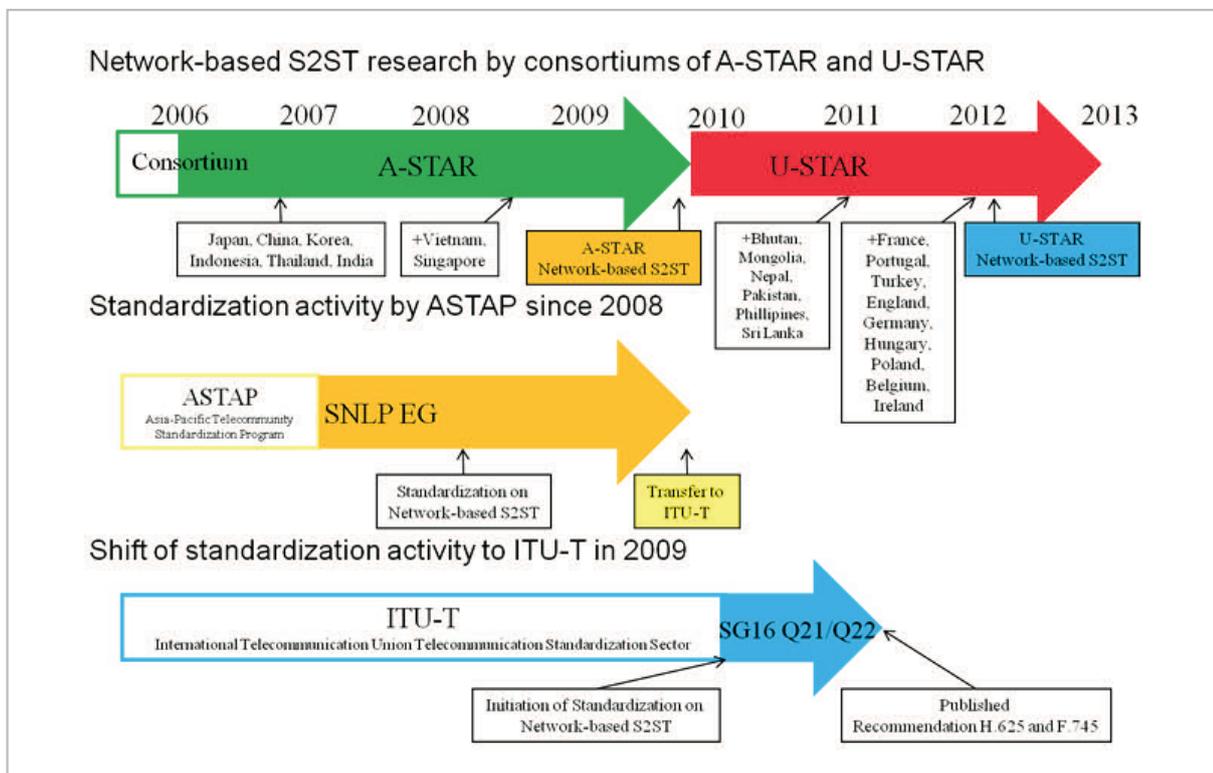


図3 国際共同コンソーシアムの拡大と通信プロトコル国際標準化活動の拡大

し拡大していく過程と、その過程において行われた標準化活動の ASTAP から ITU-T への移行を示す。

3.2 ITU-T におけるネットワーク型音声翻訳通信プロトコルの国際標準化

NICT は U-STAR の一員として、ネットワーク型音声翻訳の通信プロトコルの標準化を始動し、ITU-T の SG16、WP2、Q21/22 において「ネットワーク型音声翻訳の機能の使用：F.S2STreqs」と「ネットワーク型音声翻訳のシステム構成の仕様：H.S2Starch」という 2 種類の勧告書（表 1）の編集者を務めた。2010 年 10 月 ITU-T にて勧告案が承認され、ITU-T 勧告書 F.745 および H.625 として国際標準技術と認められ、これを

以って音声翻訳モジュールはネットワーク越しに世界中に繋ぐ事ができる。

3.3 U-STAR によるロンドンオリンピック向け音声翻訳実証実験計画

U-STAR コンソーシアムは世界にその活動を広げ、現在 23 カ国から 26 機関の加盟国をもつ共同体となった（2012 年 6 月現在）。U-STAR でサポートしている言語も 23 言語となり、世界人口の約 95.4% の話す言語を網羅している。図 4 は、U-STAR 加盟国が研究開発している言語を第一公用語として用いている地域を示す。図 5 は、U-STAR 加盟機関の一覧を示す。

U-STAR 加盟国の研究協力の下、NICT では ITU-T 勧告書 F.745、H.625 で標準化されたプロ

表 1 ネットワーク型音声翻訳プロトコルの標準化された ITU-T 勧告書

	ITU-T 勧告 F.745	ITU-T 勧告 H.625
表題	ネットワーク型音声翻訳の機能の仕様	ネットワーク型音声翻訳のシステム構成の仕様
対象範囲	音声翻訳サービスに必要な機能モジュール（音声認識、機械翻訳及び音声合成）をネットワーク接続するための要求条件、アーキテクチャ等を規定 http://www.itu.int/rec/T-REC-F.745-201010-I	機能モジュール間での通信を実現するためのインタフェース、プロトコル及びデータフォーマットを規定 http://www.itu.int/rec/T-REC-H.625-201010-I



図 4 U-STAR 加盟国とサポート言語地図

 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-Khwarizmi 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-LIMSD), France	 Institute of Systems and Computer Engineering - Research and Development in Lisbon, (INESC-ID), Portugal	 Polish-Japanese Institute of Information Technology, (PJIIT), 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				

図 5 U-STAR 加盟機関の一覧

トコルに基づき、U-STAR の各機関の音声翻訳サーバとスマートフォン上に開発したネットワーク音声翻訳アプリケーションを接続し、音声翻訳の実証実験を行っている。開発された音声翻訳アプリケーションは、対面式で対話をする単体デバイス用アプリと、対面でも離れたところでも会話できる複数デバイス用アプリの2種類である。U-STAR は、2012年7月に、これら iPhone* 上の音声翻訳用アプリケーションを英国ロンドンにて開催される U-STAR ワークショップにて公式リリースし [5]、ロンドンオリンピック 2012 を皮切りにネットワーク型音声翻訳システムの実証実験を開始した。

4 おわりに

ITU-T 標準のネットワーク型音声翻訳プロト

コルを用いることで、世界中に分散する音声翻訳モジュールを接続することができ、より多くの言語の音声翻訳を可能とする枠組みを構築した。今後、U-STAR を通して、さらに多くの世界中の研究機関が自国語の音声翻訳技術の研究に取り組み、音声翻訳ネットワークを介して実環境での実証実験を重ね、音声翻訳の性能改善を行っていく。さらに、各加盟機関が音声翻訳技術を企業に移転することにより、民間による事業化を加速させることができる。我々は、ネットワーク型音声翻訳を用いて、人類の長い間の夢であった言語の壁を越えたコミュニケーションを実現し、国際社会に大きく貢献する。

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

参考文献

- 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), Architectural Requirements 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>

(平成 24 年 6 月 14 日 採録)



ほり ちおり
堀 智織

ユニバーサルコミュニケーション研究所
音声コミュニケーション研究室
主任研究員
博士 (学術)
音声認識、音声翻訳、音声対話技術
chiori.hori@nict.go.jp