Network-based Speech-to-Speech Translation Technology

Chiori Hori
Spoken Language Communication Laboratory
MASTAR Project
National Institute of Information and Communications Technology
The National Institute of Information and Communications Technology (NICT)

NICT is in charge of deciding and maintaining Japan Standard Time.
Network Security Research Institute
Applied Electromagnetic Research Institute
Advanced ICT Research Institute
Wireless Network Research Institute
Photonic Network Research Institute
Universal Communication Research Institute - Spoken Language Communication laboratory
Research Target

1. Spoken language communication technologies
   - human and machine
     -> Spoken Dialog system
   - humans who speak different languages
     -> Speech-to-Speech Translation system

2. Rich transcription technologies
   for speech/video archives
   -> Information Retrieval system
Many different languages in the world
Overcoming the language barrier is a long-held dream of mankind.

Speech-to-Speech Translation technology

Breaking the language barrier

Speech-to-Speech Translation (S2ST)

Japanese 「私は学校に行く」

Convert to Japanese phoneme sequence "w", "a", "t".

Convert to word sequence using lexicon and grammar

Convert to English word sequence
「私は」⇒ “I”
「学校に」⇒ “to school”
「行く」⇒ “go”

Reorder word sequences according to English grammar
“I” “to school” “go”

Select appropriate waveform for English text

Corpora
History of S2ST research

1986
Read Speech
- Syntactically correct
- Clear utterance
- Limited domain “Conference Registration”

1992
Daily Conversation
- Standard expression
- Unclear utterance
- Limited domain “Hotel Reservation”

1999
Wider and Real Domain
- Wider and real domain “International Travel”
- Realistic expressions
- Noisy speech
- J-E, J-C speech translation

2006
+ More languages for translation

2008
NICT

Fundamental technologies

Rule-based Technology
Hand-made

Corpus-based Technology
Large scale corpus + Machine learning

ATR NICT
Big projects in the US and EC
Network-based S2ST

- Realizing S2ST for multi languages in the world, in a collaboration with the international society


S2ST by C-STAR, ATR (JP), CMU (US), ETRI (KR), KIT (DE), ITC/IRST (IT), CLIPS (FR)

Speech translation projects (TIDES, GALE, TC-STAR, QUERO)

- Connecting ASR, MT, and TTS modules existing in the world, via Network.

- Definition of communication protocol and data format between modules is needed.
Architecture of Network-based S2ST system

- **Transmitter**
  - Speaker of Language A
  - Digitalization of speech signals

- **Network**
  - Communication between users who speak different languages

- **Multiple Receivers**
  - Speaker of Language A
  - Digitalization of speech signals

- **Services**
  - ASR
  - MT
  - TTS
  - Sign Recognition
  - DM, etc.

- **MC client**
  - Speaker of Language A
  - Digitalization of speech signals

- **MC server**
  - ASR server
    - Conversion from speech signal to text in Language A
  - MT server
    - Conversion from text in Language A to text in Language B
  - TTS server
    - Conversion from text in Language B to speech signal
  - ASR server
    - Conversion from speech signal to text in Language B
  - MT server
    - Conversion from text in Language B to text in Language A
  - TTS server
    - Conversion from text in Language A to speech signal
MCML

- Modality Conversion Markup Language
- XML schema
- MCML includes information for communication among multiple persons who use different modalities
Conversation of multiple Language speakers

English meaning: “Where was the most interesting place you've visited?”
Multiparty conversation system
Network-based S2ST research by ASTAP consortium

Standardization activity by ASTAP since 2008

Shift of standardization activity to ITU-T in 2009
Activity start for standardization of Network-based S2ST at ITU-T SG16  
Session period: October, 2009 to the present  
NICT is the editor for S2ST standardization at ITU-T SG16, WP2 Q21/22  

<table>
<thead>
<tr>
<th>Document</th>
<th>Title</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.745</td>
<td>Functional Requirements for Network-based S2ST</td>
<td>- Definition of Network-based S2ST</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Functions and service requirements of network-based S2ST</td>
</tr>
<tr>
<td>H.625</td>
<td>Architectural Requirements for Network-based S2ST</td>
<td>- Requirements of S2ST architecture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Definition of interface for Network-based S2ST</td>
</tr>
</tbody>
</table>

Not only language conversion but also the potentially added module like sign language are taken into account:  
S2ST -> Modality conversion
Functional requirements for network-based speech-to-speech translation services

Recommendation ITU-T F.745
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"

targetNamespace="http://www.itu.int/xml-name-space/itu-t/H.645/MCML.xsd"

elementFormDefault="qualified" attributeFormDefault="unqualified">

<xs:element name="MCML">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="User"/>
      <xs:element ref="Server" maxOccurs="unbounded"/>
      <xs:element ref="History" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="Version" type="xs:string" use="required"/>
  </xs:complexType>
</xs:element>

...
Field Experiment by VoiceTra

Description

You can translate the contents of your spoken words to a foreign language using the application developed by National Institute of Information and Communications Technology (NICT).

The target conversation content for the application is text-related.

NICT Web Site: VoiceTra Speech to Speech Translator by NICT Support

What’s New in Version 3.4.0

- [Feature Additions]
  - New translation for ricerca (Italian) andره صناعات (Arabic)
  - New translation for 購物 (Japanese)

- [Functionality Improvement]
  - Speaking can now be performed by tapping the screen.

http://www.youtube.com/watch?v=v6ULvxsPh54&feature=player_detailpage
Further expansion of Asian languages and plus European languages

Growth of multi-lingual S2ST network