Proceedings of IWSLT 2009, Tokyo - Japan

ISCA Archive
http://www.isca-speech.org/archive

Network-based Speech-to-Speech Translation

Chiori Hori, Sakriani Sakti, Michael Paul, Noriyuki Kimura,
Yutaka Ashikari, Ryosuke Isotani, Eiichiro Sumita, Satoshi Nakamura

Knowledge Creating Communication Research Center, MASTAR project
National Institute of Information and Communications Technology
{chiori.hori, sakriani.sakti, michael.paul, noriyuki.kimura, ryosuke.isotani, eiichiro.sumita, satoshi.nakamura}@nict.go.jp

Abstract

This demo shows the network-based speech-to-speech translation system. The system was designed to perform real-time, location-free, multi-party translation between speakers of different languages. The spoken language modules: automatic speech recognition (ASR), machine translation (MT), and text-to-speech synthesis (TTS), are connected through Web servers that can be accessed via client applications worldwide. In this demo, we will show the multi-party speech-to-speech translation of Japanese, Chinese, Indonesian, Vietnamese, and English, provided by the NICT server. These speech-to-speech modules have been developed by NICT as a part of A-STAR (Asian Speech Translation Advanced Research) consortium project[1].

1. Network-based Speech-to-Speech Translation Systems

1.1. Architecture of Network-based S2ST

Figure 1 illustrates the overall structure of Network-based speech-to-speech translation system. This system is composed of the following components:

- Spoken language technology servers

The spoken language technologies, including ASR, MT, and TTS engines, were provided by NICT through Web servers.

- Speech Translation Markup Language (STML) servlet

All data exchanges among client users and spoken language technology servers are managed through a Web service designed by NICT, the so-called STML servlet. It follows a standard protocol, namely, STML.

- Client application

for The client applications are implemented on a handheld mobile terminal device, which allows portable speech-to-speech translation. It was developed by NICT and supports both speech and video interaction between client users.

- Communication server

A communication server, also provided by NICT, is used to relay the speech results from one user to all other users in order to enable them to perform a multiparty conversation.

[1] A-STAR consortium are consists of following members: NICT (Japan), ETRI (Korea), CASIA (China), NECTEC (Thailand), BPPT (Indonesia), CDAC (India), IOIT (Vietnam), and I^{2}R (Singapore) [1].

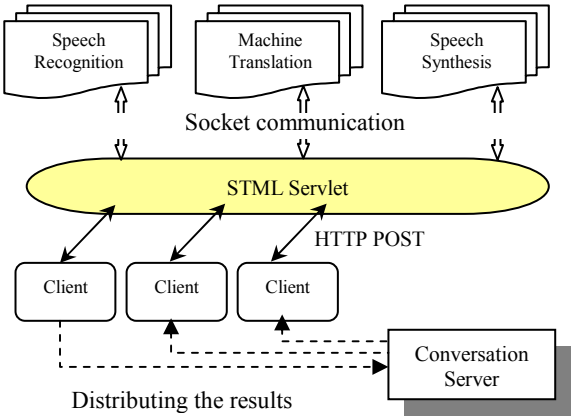

Fig. 1. Architecture of client-server interaction.

1.2. Client device

The client applications are implemented on a handheld mobile terminal device (Sony VAIO-U) shown in Fig. 4, which allows portable speech-to-speech translation. The device is 150-mm wide, 95-mm high, and 32-mm thick.

Fig. 2. The client application on a hand-held terminal device.

2. References

[1] Sakriani Sakti, Noriyuki Kimura, Michael Paul, Chiori Hori, Eiichiro Sumita, Satoshi Nakamura, Jun Park, Chai Wutiwiwatchai, Bo Xu, Hammam Riza, Karunesh Arora, Chi Mai Luong, Haizhou Li, "The Asian Network-based Speech-to-Speech Translation System," to appear in Proc. ASRU2009.