BBN TransTalk: A Portable Android-Based Interactive Speech-to-Speech Translation System

Sankaranarayanan Ananthakrishnan, Aaron Challenner, Stavros Tsakalidis, Shirin Saleem, David Stallard, Chia-lin Kao, Fred Choi, Ralf Meermeier, Mark Rawls, Jacob Devlin, Kriste Krstovski, Rohit Prasad, Prem Natarajan

Speech, Language and Multimedia Technologies
Raytheon BBN Technologies
10 Moulton Street
Cambridge, MA 02138, U.S.A.

Abstract

Robust, interactive speech-to-speech (S2S) translation is a technology highly sought after by various government and industrial agencies for a wide range of applications. These include but are not limited to tactical military missions, border protection, and medical evaluation and diagnosis in clinics/hospitals with a significant non-native clientele.

To address this challenge, BBN has developed *TransTalk*, a highly portable, scalable, robust, interactive end-to-end S2S system that:

- Performs large-vocabulary automatic speech recognition (ASR) in the source language.
- Incorporates large-vocabulary phrase-based statistical machine translation (SMT)
- Generates an audible rendition of translated text in the target language using high-fidelity text-to-speech synthesis (TTS).

Perhaps the most remarkable aspect of TransTalk is that it runs entirely on a single off-the-shelf Android smart phone. The system is self-contained and does not rely on any external server for processing, unlike other cloud-based translation services such as Google's Translate app. It is thus suited for deployment anywhere in the world regardless of the local wireless infrastructure.

Summarized below are some of the numerous engineering and research breakthroughs that enabled the TransTalk system.

- *Multi-pass decoding* using our patented two-pass search strategy [1].
- *Fast Gaussian Computation* (FGC) reduces the number of evaluated Gaussians per frame by a factor of 4. Further speedup was obtained by jointly quantizing the Gaussian means and variances to 8 bits for precomputing the Gaussian distance for each frame, before the search.
- A low-latency, "streaming" version of *online speaker adaptation* [2] for acoustic models.
- *On-the-fly loading* of acoustic, language, and translation models (AM, LM, and TM) from disk for memory conservation in ASR and SMT.
- *Boosted phrase table* [3] for reducing the number of irrelevant TM translations.
- *Precomputed LM probabilities* for phrase-internal target words.
- Dynamic *on-line LM adaptation* [4].
- *Phrase alignment confidence* [5] for improved target phrase selection.
- Ability for the user to barge in, or abort, the translation pipeline at any stage and return the device to a ready state.
- In *dual-phone mode* using Bluetooth, each speaker has their own phone and can be separated by up to 30 feet.

Currently, the Transtalk system supports full two-way interaction between English speakers and Iraqi Arabic, Dari, and Pashto speakers. It covers a wide variety of domains ranging from force protection to medical evaluation.

References

[1] L. Nguyen, R. Schwartz, 1997. Efficient 2-pass n-best decoder.In *DARPA Speech Recognition Workshop*, pp. 167-170.

[2] D. Liu, D. Kiecza, A. Srivastava, F. Kubala (2005), "Online Speaker Adaptation and Tracking for Real-Time Speech Recognition", in Proceedings of Interspeech, pp. 281-284

[3] S. Ananthakrishnan, R. Prasad, P. Natarajan (2010), "An Unsupervised Boosting Technique for Refining Word Alignment", in *Proceedings of IEEE SLT Workshop*, pp. 177-182, Berkeley, CA

[4] S. Ananthakrishnan, R. Prasad, P. Natarajan (2011), "On-line Language Model Biasing for Statistical Machine Translation", in *Proceedings of ACL*, pp. 445-448, Portland, OR

[5] S. Ananthakrishnan, R. Prasad, P. Natarajan (2010), "Phrase Alignment Confidence for Statistical Machine Translation", in *Proceedings of Interspeech*, pp. 2878-2881, Makuhari, Japan