Speech Translation Systems as a Solution for a Wireless Earpiece

Nicholas Ruiz, Andrew Ochoa, Jainam Shah, William Goethels, Sergio DelRio Diaz
Waverly Labs, Brooklyn, NY, USA
{nick,andrew,jainam,william,sergio}@waverlylabs.com

Abstract

The advances of deep learning approaches in automatic speech recognition (ASR) and machine translation (MT) have allowed for levels of accuracy that move speech translation closer to being a commercially viable alternative interpretation solution. In addition, recent improvements in micro-electronic mechanical systems, microphone arrays, speech processing software, and wireless technology have enabled speech recognition software to capture higher quality speech input from wireless earpiece products. With this in mind, we introduce and present a wearable speech translation tool called Pilot, which uses these systems to translate language spoken within the proximity of a user wearing the wireless earpiece.

1 What is Pilot?

The Pilot Translating Earpiece is a sophisticated earbud which uses dual microphones and custom noise cancelling algorithms to produce clear speech before it is passed through our mobile app and to our speech translation engine in the cloud. It relays speech translation very quickly with minimal latency. Pilot consists of two translation earbuds that pair with custom speech translation software for Android or iOS. Pilot allows consumers to share their secondary earbud with a conversation partner for face-to-face simultaneous speech translation and currently supports 15 languages[1].

[1]Arabic, Chinese (Mandarin, Cantonese), English, French, German, Hindi, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Greek, Turkish, and Polish ASR and MT.

2 How Does Pilot Work?

Pilot operates in two modes: *Converse* and *Listen*. As the primary use case, *Converse* mode allows multi-party conversations with transcriptions logged in the app. In a one-on-one conversation, users can share their secondary earbud with a partner and quickly pair it with the partner's phone. Currently in beta, *Listen* mode adapts the microphone firmware settings to pick up ambient sound and performs far-field ASR and MT. Pilot uses several speech translation paradigms, depending on the language pair, either by running ASR and MT sequentially, or as tightly coupled speech translation[2]. Translations are primarily run on the server, while the app is responsible for routing the audio to and from the earpiece.

Practical challenges *Bluetooth*: Conventionally, Android and iOS devices are limited to one microphone connection at a time. Although routing the partner's earbud recordings through the same phone is possible, it requires low-level kernel programming to implement. Our team will resolve this issue in a future release. *Microphone pick-up*: Occasionally a conversation partner's speech can be picked up by the user's earbud, and vice-versa. While digital signal processing can eliminate some of this effect, the position, distance, and power of the speech must be taken into account.

As the provider of one of the first translation wearables to market, we are eager to how learn translation technology affects situational dialogue without an interpreter present. While our first version pieces maturing technologies together, we are working on improving the user experience by minimizing user's dependence on their phone's screen.

[2]Speech synthesis is currently not informed by ASR or MT.

Pérez-Ortiz, Sánchez-Martínez, Esplà-Gomis, Popović, Rico, Martins, Van den Bogaert, Forcada (eds.)
Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 361
Alacant, Spain, May 2018.