The MUTE Silent Speech Recognition System

Geoffrey S. Meltzner¹, James T. Heaton², Yunbin Deng¹

¹ BAE Systems, Inc. Burlington, MA, USA
² Massachusetts General Hospital Department of Surgery, Boston, MA, USA

geoffrey.meltzner@baesystems.com, james.heaton@mgh.harvard.edu, yunbin.deng@baesystems.com

Abstract

sEMG based silent speech recognition has become a desirable communication modality because it has the potential to provide natural, covert, hands-free communication in acoustically challenging environments. To enable this capability, we have developed a portable, self-contained, Android based Mouthed-speech Understanding and Transcription Engine (MUTE) system. To demonstrate the MUTE system’s ability to recognize a continuous vocabulary of 210 words we propose to conduct map task based demonstration, in which a MUTE user guides a “listener” around a schematized map. The listener draws out a map based on the received instructions; a comparison of the original and drawn map then illustrates the MUTE system’s recognition performance.

1. Background

Non-acoustic-based automatic speech recognition has the potential to mitigate two significant weaknesses of standard, acoustic ASR: (1) severe performance degradation in the presence of ambient noise and (2) a limited ability to maintain privacy/secrecy because of the requirement of using audible speech. Recent non-acoustic ASR studies have investigated alternative modalities, such as ultrasound [1] or surface electromyography (sEMG) [2-9] that can capture sufficient speech information while overcoming the aforementioned deficiencies of acoustic ASR systems.

sEMG-based speech recognition, also known as subvocal speech recognition, operates on signals recorded from a set of sEMG sensors that are strategically located on the neck and face surface to measure muscle activity associated with speech production. Because the signals directly measure articulatory muscle activity there is no need for acoustic excitation of the vocal tract, making it possible to recognize silent, mouthed speech. Moreover, because sEMG signals are decoupled from acoustic signals, they are immune to acoustic noise corruption.

2. The MUTE system

Our research in this domain [2],[10],[13] (and paper at INTERSPEECH 2013 titled “Augmenting sEMG-Based Speech Recognition by Tracking Lingual-Palatal Contact”), has culminated in the development of the MUTE (Mouthed-speech Understanding and Transcription Engine) silent speech recognition system. The system, as shown in Figure 1, consists of four components: (1) a set of 4 pairs of wireless sEMG electrodes custom designed to be placed on the face and neck which connect to (2) a wireless transmitter module. The wireless transmitter communicates with the portable wireless receiver module (3). The receiver module transmits the raw sEMG data via a USB connection to a portable Android device (4), which runs the MUTE recognition software. The MUTE recognition software currently takes the form of a “Push-To-Talk” silent chat app that communicates over a WiFi network with other MUTE systems. Figure 2 provides a screen capture of the MUTE app to illustrate its current form.

Figure 1. The MUTE system is comprised of (1) four pairs of customized sEMG sensors; (2) a wireless transmitter module; (3) a wireless receiver module; and (4) an Android mobile device running the MUTE recognition app.

Figure 2. Screen capture of the MUTE app.

As currently configured, MUTE is an entirely self-contained, portable system that enables hands-free silent communication in real world environments. Figure 3 shows the MUTE system being demonstrated in an outdoor environment while the user is situated in non-ideal pose.

3. MUTE Demonstration

To illustrate MUTE’s current capabilities and recognition performance, we propose to conduct a demonstration based around a map task. In this scenario, a user outfitted with a complete MUTE system, called the “speaker” will silently communicate with another user, (the “listener”) outfitted with an Android tablet with the MUTE app running in a receive-
only mode. The two MUTE systems will be linked over a local WiFi network. The speaker will guide the listener through one of five possible schematized map (see ) instructing the listener to “move” a certain distance and direction to a set of different landmarks. At each landmark the listener will be instructed to “perform” a one of a limited set of tasks. Based on the received instructions, the listen will draw out a map and compare it to the original map. The discrepancies between the two provide an immediate visual measure of the MUTE system’s recognition capabilities.

4. Acknowledgements

The authors would like to thank Carlo J. De Luca, Gianluca De Luca, Don Gilmore, and Serge Roy of Delsys, Inc. for for developing and providing the sEMG sensors and related hardware. This study was sponsored by the United States Defense Advanced Research Projects Agency (DARPA) Information Innovation Office (I2O) Program, “Advanced Speech Encoding.” The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of DARPA or the U.S. Government.

Figure 3. The MUTE system being demonstrated in an outdoor environment. Note that in this case, the MUTE software is running on an Acer Iconia A500 tablet, rather than the Samsung Galaxy S2 mobile phone shown in Figure 1.

Figure 4. An example of a schematized map to be used in the demonstration. Five possible maps will be available to show the flexibility of the MUTE recognition system.
5. References


