Small Vocabulary Communication and Control Using Surface Electromyography in an Acoustically Noisy Environment

Bradley J. Betts[1] and Charles Jorgensen[2]

Neuro-Engineering Laboratory
NASA Ames Research Center, M/S 269-1
Moffett Field, CA 94035-1000

[1]*QSS Group, Inc., bbetts@email.arc.nasa.gov*
[2]*NASA, cjorgensen@email.arc.nasa.gov*

1. Extended Abstract

Speech intelligibility can be severely degraded by high levels of acoustic noise. Researchers have developed a variety of techniques to minimize the impact of noise, ranging from adaptive noise cancellation to throat microphones. Increasingly, researchers are experimenting with the measurement and analysis of bioelectric signals associated with speech in an effort to further minimize—or even completely eliminate—the degrading effects of acoustic noise. Such techniques, either on their own or fused with other modalities, hold promise for improving human communication and human-computer interaction.

The bioelectric technique used in this research is electromyography, the study of muscle function through its electrical properties. Electrical activity associated with speech can be detected by non-invasive surface sensors mounted in the region of the face and neck. Sensing of this type is not directly interfered with by acoustic noise (although indirect effects, akin to the Lombard effect, require further study).

First responders are an example of a class of users that stand to benefit from reliable communication in acoustically harsh environments. For example, sirens, engines, and saws all add noise to a typical firefighting scene, as does the breathing apparatus a firefighter wears. This work was motivated in part by a desire to see whether EMG-based speech recognition could alleviate these effects (and is part of a broader effort at NASA investigating the physiology, signal processing, and applications of EMG-based speech recognition).

In the portion of the research reported here, EMG data were collected from a male subject under laboratory conditions. Data samples were used to train a neural network classifier and to test the generalizability of the network. The trained network was then inserted into a real-time communication and control system while the subject was exposed to approximately 80 dB acoustic noise. Isolated phrases recognized from the EMG signal can be communicated to a cellular phone or used to control a robotic platform (see Figure 1).

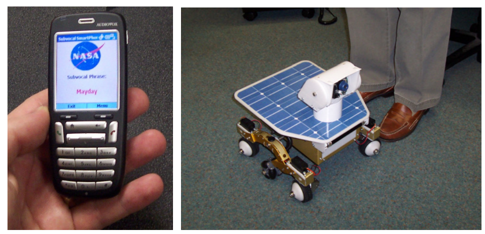

Figure 1. Communication and control output modalities: cellular phone and robotic platform (Carnegie Mellon University Personal Exploration Rover).

For each phrase used by the system, 150 samples were collected. Of these, 120 were used to train a neural network while 30 randomly-chosen samples were set aside for generalizability testing. A single differential channel of EMG data was recorded from sensors on the subject's neck (see Figure 2) using a Synamp amplifier (Neuroscan Compumedics; El Paso, TX). The signal was sampled at 10 kHz and band-pass filtered from 10 Hz to 2 kHz. A 60 Hz notch filter was used to reduce line noise. During both data collection and subsequent real-time testing, the subject was wearing a self-contained breathing apparatus (SCBA) on loan from the Moffett Field Fire Department.

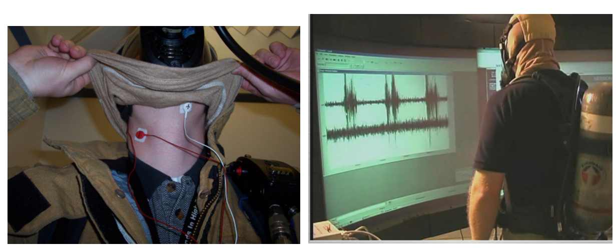

Figure 2. Photos showing sensor placement and EMG data collection. The subject has peeled his hood back to reveal the sensors (a third sensor, placed behind the subject's right ear, is used for grounding).

The signal processing and pattern recognition techniques used in this work are similar to ones we have previously reported [1, 2]. Activity detection is used to segment phrases out of the EMG data. Feature extraction is then performed using wavelet transforms, with the labeled features fed into a neural network for training.

For six phrases (“Evacuate”, “Mayday”, “Man-trap”, “Fire Safe”, “Zero”, “One”) a confusion matrix was generated from the samples in the test set. The lower bound on the various 95% correct classification confidence intervals was 75%; the upper bound was 100%. In real-time testing, shown in Figure 3, the subject was seated in a laboratory setting. Common sounds encountered in firefighting were played through speakers at approximately 80 dB to make normal speech communication difficult. Recognized phrases could be communicated to a cellular phone (where they were displayed visually on the screen) and used to control a robotic device (by mapping phrases to movements such as forward, turn left, etc.).

Figure 2. Real-time EMG-based communication to a cellular phone.

We observed no change in recognition rates due to acoustic noise or to the use of an SCBA. These observations lead us to conclude that EMG-based speech recognition shows promise for first responder environments. Many challenges need to be overcome, however, before a viable system can be fielded. This includes expanding the size of the recognized vocabulary, miniaturizing and hardening equipment, and increasing the recognition rate. Future work will look to address these challenges. It will also look at the viability of a multi-modality approach to communication and control, for example using EMG-based speech to augment conventional speech systems in a first responder environment.

2. Selected References

[1] C. Jorgensen and K. Binsted, "Web Browser Control Using EMG Based Sub Vocal Speech Recognition," *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, IEEE, 2005, pp. 294c.1-294c.8.

[2] C. Jorgensen, D. D. Lee, and S. Agabon, "Sub auditory speech recognition based on EMG signals," *Proceedings of the International Joint Conference on Neural Networks*, IEEE, 2003, pp. 3128-3133.