Development of A Visual Speech Recognition Program

Arun Raghavan[1], Gavriel D. Kohlberg[2], Joseph T. Breen[2], Ravi N. Samy[2]

[1]University of Cincinnati Medical School,[2] Division of Otology/ Neurotology, Department of Otolaryngology, University of Cincinnati / Cincinnati Children's

Introduction:

Hearing loss is a major societal burden associated with social isolation and a decreased quality of life. Diminished speech perception is a major barrier to communication for affected individuals, especially in environments with background noise. We have developed a visual speech recognition program (VSRP) that uses artificial intelligence to translate articulatory lip movements into speech text. The VSRP can potentially augment a subject's speech perception in noise by presenting a transcript of the speech to the subject in conjunction with the audible speech.

Hypothesis:

Individuals will achieve greater speech perception in noise with the aid of a VSRP.

Methods:

A dual video/infrared camera (Microsoft Kinect) was used to continuously track 35 points around lips during speech in real-time. A VSRP was developed to identify 42 sentences from the BKB SIN speech-in-noise test using the recorded articulations. Seven normal-hearing individuals were recruited. Participants were situated in a sound-isolated audio booth such that they could see the speaker through a window. Each participant was asked to repeat sentences presented by a speaker at 50dB with varying degrees of background noise. Speech perception accuracy was evaluated for each subject on 40 sentences from BKB-SIN without the VSRP and 42 sentences with the VSRP.

Results:

Participants achieved an average accuracy of 97.04% in low noise conditions (noise <70dB) and 37.94% in high noise conditions (noise ≥ 70dB). With the VSRP, average accuracies were 100% in low noise conditions and 73.50% in high noise conditions. The VSRP led to an average increase in accuracy over all listening conditions of 22.47±3.37 (%±SE, $p = 0.0006$). In the high noise condition, the average increase in accuracy was 35.56±4.78 (%±SE, $p = 0.0001$).

Conclusion:

A statistically significant increase in accuracy was observed with the use of a real-time VSRP for normal hearing listeners. Application of VSRPs could potentially increase speech perception in noise for individuals with hearing loss. Further evaluation is needed to demonstrate the functionality of this system for patients with hearing loss.

Acknowledgements:

This study was supported in part by a NIH grant T35DK06044.