SINGLE-CHANNEL VIBROTACTILE FEEDBACK FOR VOICING ENHANCEMENT IN TRAINED AND UNTRAINED PERCEIVERS

David Marino [*1,2], **Hannah Elbaggari** [†1,2], **Tzu Hsu Chu** [‡1,2], **Bryan Gick** [§2], **and Karon MacLean** [¶1]

[1]University of British Columbia, Department of Linguistics
[2]University of British Columbia, Department of Computer Science

1   Introduction

We evaluate the feasibility of using a single-channel vibrator to enhance the intelligibility of speech in acoustically noisy environments. Speech intelligibility can be enhanced by integrating information from other modalities, e.g., vision [1]. Nonetheless, there exists circumstances where shared visual attention is hard to establish, or in-person contact is infeasible (e.g., in a noisy collaborative environment such as an industrial shop floor). Previous research has established that direct manual touch can enhance speech intelligibility in noisy environments [2]. Touch is an ideal modality to address this problem.

Tactile aids have been used to enhance speech intelligibility through multichannel vibrotactile stimulation in clinical populations (including individuals who are hard of hearing) [3] [4] [5]. But little work has been done to address the feasibility of these devices for non-clinical populations, particularly using single-channel systems. Establishing the efficacy of a single-channel vibrotactile device is desirable because many people already own a single-channel vibrotactile device (such as a cellphone), and it could therefore be easily implemented in such everyday objects.

We informed the design of the system from a pilot study and a historically popular method of tactile speech enhancement, the Tadoma method, where listeners place their hands on a speaker's face and throat [6] [2]. Our device provides vibrotactile stimulation similar to the laryngeal vibrations felt in the Tadoma method by vibrating to the amplitude envelope of voiced speech. A pilot study with untrained perceivers was conducted to establish the ground truth that vibrotactile stimulation can enhance the intelligibility of speech as well as to inform the design of the system. Participants were tasked with identifying the content of speech in noise with and without the aid of a vibrator. Significant variance was attributable to vibration style ($F_{(2,48)}=4.21$, p=0.01), as well as phonemic contrast ($F_{(2,48)}=15.395$, p=$5.83 \cdot 10^{-6}$), but not vibrator placement. In the current study, a training phase was added (per results of [7]) and unpromising conditions from the pilot were eliminated. We demonstrate that vibrotactile feedback increases the accuracy scores of participants when discerning the content of speech in noise.

[*]damarino@cs.ubc.ca
[†]hre@cs.ubc.ca
[‡]sophychu@cs.ubc.ca
[§]gick@mail.ubc.ca
[¶]maclean@cs.ubc.ca

2   Methods

2.1   Participants

20 participants were recruited from the University of British Columbia's Linguistics human subject pool. They were compensated with either course credit or $10 for their time.

2.2   Materials and Procedure

Stimuli

180 minimal pairs were utilized and randomized within testing trials. A subset of 120 words (60 per trial block) were used. Stimuli were recorded using the voices of male and female identifying English native-speakers with a DR40 TASCAM hand held linear PCM digital recorder. The volume between audio tokens was normalized before use in the experiment proper.

Procedure

Vibrations were administered through a Tectonic Elements Audio (TEAX12C02-8/RH) linear resonant acutator that was held between the index finger and thumb as to target the sensitive glabrous skin of the hands. The vibrotactile waveforms were procedurally generated from speech during the experiment. The vibrations were designed to mimic the laryngeal vibrations normally felt during voiced speech: they were only present on voicing, and the amplitude of the vibrations were coupled to the amplitude envelope of the speech. Auditory speech and noise were delivered through AKG over-the-ear headphones.
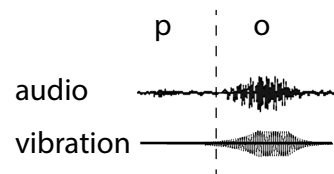

Figure 1: Above: example of how vibrations are generated from auditory speech. /p/ is unvoiced so there is no vibration present.

For each trial, we displayed a minimal pair on the computer screen accompanied by the audio of one of the words through the headphones (and optionally vibrations). Participants were then tasked with selecting the correct word. The experiment consisted of the following phases:

(1) *Training:* The training phase was intended to help participants get a feel for the speech-to-vibration mappings in an easy low-stakes environment. They were tasked with

making binary selections between minimal pairs over a low-volume babble track, with the aid of the vibrator. Each new word token was introduced with a 3 second visual countdown. If they chose incorrectly, they were prompted to try again. If they chose correctly, the task would continue for a total of 12 training trials until the next phase.

(2) *Calibration:* The calibration phase was similar to training phase, with removal of the countdown, vibrations, and correctness feedback. During calibration, we staircased the noise floor such that the participant's accuracy scores were just above chance (70% $\pm 5\%$).

(3) *Main phase:* During the main phase participants made binary selections between minimal pairs over a high-volume babble track. The main phase was blocked into a control block (where the participants held the vibrator but it did not provide vibrations) and treatment block (where the vibrator provided vibrations). Each block consisted of 60 minimal pairs, with pair order randomized, and block order counterbalanced between subjects.

3 Results

A 2-way repeated measures ANOVA was conducted to examine the effects of vibrations, and phonemic contrast on participant accuracy scores. There was a significant effect of vibration on accuracy scores at the $p<0.01$ level [$F(1,19)=14.69$ $p=0.0012$] and phonemic contrast on accuracy scores at the $p<0.01$ level [$F(1,38)=14.9$ $p=1.66\cdot10^{-5}$]. We found no significant interaction effect between phonemic contrast and vibrations. Participants performed better with the vibrator than without, with an 8.91% difference of means between the control and vibrotactile conditions.
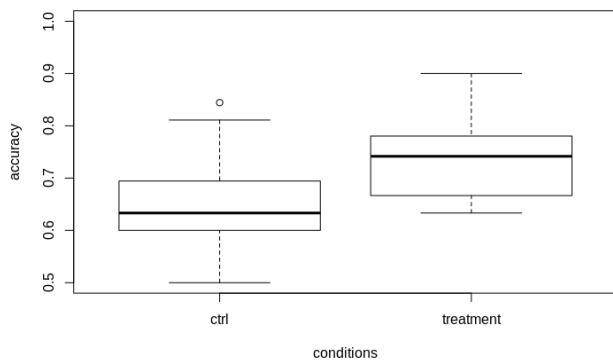
Figure 2: participant accuracy scores compared to control (no vibrations) and treatment (with vibrations)

4 Discussion

We demonstrated that single channel vibrotactile stimulation can effectively enhance the intelligibility of speech in noisy environments. The results show potential in integrating the system into an everyday object such as a cellphone. The underpinning cognitive mechanisms of vibrotactile enhance-
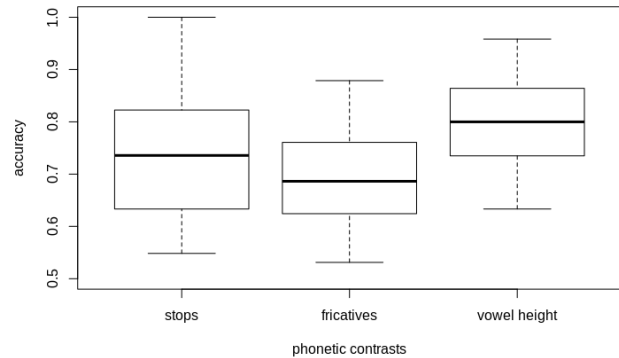
Figure 3: comparison of phonetic contrasts and participant accuracy scores when administering vibrotactile stimulation.

ment remains unanswered for a follow up study. Accuracy scores could be enhanced because participants are using the vibrotactile signal to supplement linguistic content. However, in unstructured qualitative post-study interviews, many participants claimed that the vibrations helped more to discern when to pay attention to the relevant speech in noise. Thus, it is very likely that the vibrotactile signal could serve as an attentional aid, and need be only minimally motivated by the form of the speech signal.

Acknowledgments

References

[1] William H Sumby and Irwin Pollack. Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, 26(2):212–215, 1954.

[2] Bryan Gick, Kristín M Jóhannsdóttir, Diana Gibraiel, and Jeff Mühlbauer. Tactile enhancement of auditory and visual speech perception in untrained perceivers. *The Journal of the Acoustical Society of America*, 123(4):EL72–EL76, 2008.

[3] D Kimbrough Oller. Tactile aids for the hearing impaired: An overview. In *Seminars in Hearing*, volume 16, pages 289–295. Copyright© 1995 by Thieme Medical Publishers, Inc., 1995.

[4] Karyn L Galvin, Gina Mavrias, Alessandra Moore, Robert SC Cowan, Peter J Blamey, and Graeme M Clark. A comparison of tactaid ii and tactaid 7 use by adults with a profound hearing impairment. *Ear and hearing*, 20(6):471–482, 1999.

[5] Mark D Fletcher, Sean R Mills, and Tobias Goehring. Vibrotactile enhancement of speech intelligibility in multi-talker noise for simulated cochlear implant listening. *Trends in hearing*, 22:2331216518797838, 2018.

[6] Sophia Alcorn. The tadoma method. *Volta Rev*, 34(1932):195–198, 1932.

[7] Lynne E Bernstein, Silvio P Eberhardt, and Marilyn E Demorest. Single-channel vibrotactile supplements to visual perception of intonation and stress. *The Journal of the Acoustical Society of America*, 85(1):397–405, 1989.