Modeling Tactile Speech Perception with Auditory Simulation

Albert T. Lash, Janet M. Weisenberger, and Ying Xu

Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

Abstract: A useful technique for gaining insights into the perception of speech stimuli presented via vibrotactile aids is to model this perception by presenting similarly processed auditory stimuli to normal-hearing observers. It is likely that spectral, temporal, and intensive properties of the tactile system could all affect tactile speech perception. In the present experiments, tactile perceptual data were collected for a set of nonsense syllables presented via a vibrotactile vocoder to normal-hearing subjects. The same subjects were then tested with sets of altered auditory stimuli that were modified by manipulating spectral, temporal, or intensive parameters. Spectral alterations alone, achieved by reducing the number of filter channels and broadening the filter bandwidths, did not appreciably affect auditory performance, nor did intensity alterations, achieved by reducing the dynamic range and intensity quantization. Although temporal alterations, achieved by low-pass filtering the stimuli, did result in lower perceptual scores, the pattern of stimulus confusions did not correlate highly with the tactile data. Pairwise and three-way combinations of spectral, temporal, and intensive parameters yielded more promising correlations. Implications for constructing a comprehensive model of vibrotactile speech perception are discussed.

INTRODUCTION

Tactile aids can provide valuable acoustic information to persons with profound hearing impairments. However, although promising results have been shown for tactile speech perception, the training time to achieve these results is much longer than those of other sensory modalities (1), and the asymptotic performance does not approach that of normal auditory speech perception. Because the psychophysical characteristics of the tactile system differ from those of the auditory system, insights into the differences in tactile and auditory speech perception can be gained by determining how tactile resolution abilities affect the processing of stimuli presented via tactile aids.

A useful technique for modeling tactile speech perception is to measure auditory perception for stimuli whose features have been altered in ways that might reflect tactile system resolution. Other researchers (2,3) have successfully modeled speech perception with cochlear implants using this method.

Previous work in our lab (4) shows that vocoder-processed stimuli are very intelligible auditorily, suggesting that information reductions from tactile processing alone do not account for differences in performance. To create auditory stimuli that reflect tactile system resolution, cross-modal spatial, temporal and intensity resolution differences need to be considered. Spatial/spectral differences may result when the distance between elements on the tactile display is less than the 2-point resolution abilities of the tactile system. Temporal resolution in the tactile system can be five to ten times longer than that in the auditory system (5), leading to temporal differences in processing. And finally, the smaller dynamic range of the tactile system (6) can lead to intensity differences.

The present paper reports the results of single, pairwise, and three-way manipulations of all three parameters for stimuli presented via the Queen's University tactile vocoder, a 16-channel vibratory device. Spectral alterations were achieved by decreasing the number of channels and broadening the filter bandwidths, temporal alterations by low-pass filtering the vocoder processed signal, and intensive alterations by reducing the dynamic range of the speech signal to 50 dB and quantizing the signal into five levels. These spectral, temporal, and intensive alterations were implemented as single parameter manipulations, pairwise alterations (spectral plus temporal, spectral plus intensity, and temporal plus intensity), then three-way combinations.

METHOD

Subjects: One male and 4 female subjects between the ages of 24 and 28 participated in each experiment (one male and 3 female subjects for the three-way combinations). One female subject participated in both Experiments 1 (spectral alone) and 2 (spectral plus temporal and spectral plus intensity), and one female and one male subject participated in both Experiments 2 and 3 (temporal plus intensity and three-way combinations). All other subjects were new to the study for each successive experiment. All subjects had clinically normal hearing, and had approximately four weeks of training on the Queen's vocoder to familiarize them with its operation before testing began.

Apparatus: The tactile device used was a microprocessor-based implementation of the Queen's University vocoder (7).

Acoustic stimuli are passed to a band of 16, 1/3 octave digital bandpass filters, with center frequencies between 140 Hz and 6350 Hz. The envelope energy in each channel is logarithmically compressed and used to modulate the amplitude of...
a 100 Hz square wave, which in turn drives one of the 16 magnetic solenoids mounted in a 19.5 cm linear array on the underside of the forearm. Worn in this manner, high frequency information is felt near the elbow and low frequency information is felt near the wrist.

Speech stimuli were nine nonsense speech tokens that differed in initial consonant followed by /æt/ recorded by a single female talker. For the auditory conditions, these stimuli were processed by a software simulation of the vocoder processor and presented to the subject's right ear via headphones. For auditory stimuli, channel envelopes were used to modulate either a sine wave or a 1/3 octave narrow band noise carrier at the center frequency of each channel. Spectral manipulations included 16-, 8-, and 4-channel reductions, with bandwidth increased accordingly. Temporal manipulations were achieved by lowpass filtering at 250, 100, 50, and 20 Hz. Intensity quantization and dynamic range reductions together comprised the intensity manipulation. Two-way and three-way conditions were constructed by combining these values.

**Procedure:** All tactile and auditory sessions were programmed in the CSRE 4.5 experiment generation module. Each session presented trial blocks containing the closed set of speech tokens in isolation three times each in random order, for a total of 27 trials per session. Subjects completed 20 sessions under the tactile presentation condition after familiarization training, and 5 blocks under each of the 68 auditory conditions.

**RESULTS AND DISCUSSION**

Spectral reductions alone and intensity reductions alone did not appreciably affect auditory performance. Spectral plus intensity alterations resulted in minor reductions in speech perception at the 4-channel level, but average performance for this task was still well above that of average tactile speech perception. Temporal alterations alone and spectral plus temporal alterations resulted in perceptual scores comparable to tactile performance, especially in the 100, 50, and 20 Hz lowpass filter conditions. Temporal plus intensity alteration also led to lower perceptual scores, as did the three-way combinations, especially those involving the 4-channel spectral reductions.

Pearson r correlations were calculated comparing tactile confusion matrices with auditory confusion data. All correlations were performed with diagonal values (percent correct) removed from the matrices. As expected, the high levels of performance in the spectral alone and intensity alone conditions did not yield significant correlations with the tactile data. The temporal alone reductions yielded significant positive correlations (p<.01) for two lowpass filter values (250 Hz-sines and 100 Hz-sines) of .35 and .38. Under combined spectral and temporal modifications several conditions produced significant correlations in the vicinity of .35, with one condition (4-channel, 250 Hz-sines) reaching an r of .44. Similarly, a number of significant correlations between .38 and .42 were observed for the spectral plus intensity modifications. Finally, for the three-way combinations, 6 conditions produced significant correlations in the range of .32 to .40.

Overall, it can be seen that these correlations are only moderate, and do not account for a large proportion of the variance. There are a number of possible explanations for the low correlations, including the fact that subjects were not highly trained with the tactile conditions, and might have produced higher performance levels with more training time. In addition, the present model does not take into account the physical response of the actuators on the Queen's vocoder display. It is likely that the sluggish temporal characteristics of these actuators may have further distorted the tactile signal. These issues will be explored in subsequent experiments.

**ACKNOWLEDGMENTS**

The present study was supported by a grant from the National Institutes of Health (DC00306) to the second author.

**REFERENCES**