

Tactile enhancement of auditory and visual speech perception in untrained perceivers

Bryan Gick

*Department of Linguistics, University of British Columbia, Vancouver, British Columbia, V6T 1Z4 Canada
and Haskins Laboratories, New Haven, Connecticut 06511
gick@interchange.ubc.ca*

Kristín M. Jóhannsdóttir

*Department of Linguistics, University of British Columbia, Vancouver, British Columbia V6T 1Z4 Canada
stina@mail2skier.com*

Diana Gibrael

*Center for Middle Eastern Studies, University of Chicago, Chicago, Illinois 60637
dgibrael@uchicago.edu*

Jeff Mühlbauer

*Department of Linguistics, University of British Columbia, Vancouver, British Columbia V6T 1Z4 Canada
jefmuehl@interchange.ubc.ca*

Abstract: A single pool of untrained subjects was tested for interactions across two bimodal perception conditions: audio-tactile, in which subjects heard and felt speech, and visual-tactile, in which subjects saw and felt speech. Identifications of English obstruent consonants were compared in bimodal and no-tactile baseline conditions. Results indicate that tactile information enhances speech perception by about 10 percent, regardless of which other mode (auditory or visual) is active. However, within-subject analysis indicates that individual subjects who benefit more from tactile information in one cross-modal condition tend to benefit less from tactile information in the other.

© 2008 Acoustical Society of America

PACS numbers: 43.71.Es [AL]

Date Received: October 2, 2007

Date Accepted: December 15, 2007

1. Introduction

It has long been known that both visual (McGurk and MacDonald, 1976) and tactile (Sparks *et al.*, 1978) information can influence auditory speech perception. Previous research on the effect of tactile information on speech perception has focused primarily on enhancing the communication abilities of deaf-blind individuals (Chomsky, 1986; Norton *et al.*, 1977; Reed *et al.*, 1989). Central to much of this research is the Tadoma method (Alcorn, 1932). Tadoma involves a perceiver placing his or her hand on a speaker's face in a specific position in order to gain tactile speech information. Typically, deaf-blind individuals require years of training in order to use Tadoma successfully. While a minority of previous studies have tested the use of the Tadoma method by individuals with normal hearing, speech, and visual systems (Blamey *et al.*, 1989, Reed *et al.*, 1978, 1982), the subjects of these studies were researchers who had extensively studied the Tadoma method, and who had additional training prior to undergoing the experiment.

In one of the few experiments demonstrating the influence of tactile information on speech perception in a completely untrained population, Fowler and Dekle (1991) showed that manual tactile contact with a speaker's face coupled with incongruous auditory input can elicit an audio-tactile (AT) McGurk-type illusion. However, the extent to which different modalities interact with one another in speech perception is not known. A relevant property of intermodal interaction is superadditivity, whereby a perceiver gains more benefit from information via one

modality when that information is coupled with information via another modality (e.g., McGrath and Summerfield, 1985). It is unknown, however, whether there exist complex relationships across modalities, such that a particular modality will receive greater enhancement when coupled with specific other modalities, or whether such interactions may vary across perceivers.

The present paper uses the Tadoma method to test interactions across two bimodal perception conditions for a single group of untrained perceivers. One condition involved audio-tactile (AT) speech perception in noise by perceivers unable to see the speaker, and one involved visual-tactile (VT) speech perception by perceivers when the speech was masked by noise.

2. Method

Twelve native speakers of North American English between the ages of 20 and 40 participated in the study—five male, and seven female. All subjects had normal speaking and hearing skills, and no training in linguistics, speech sciences, or any method of tactile speech perception.

Video and audio of all trials were recorded using a Sony DCR-TRV19 mini-DV camcorder. An experimenter (the third author, a female native speaker of English) read aloud from prepared lists of token utterances. All tokens on the list had the structure /aCá/, where the intervocalic consonant in each token was one of the 14 English obstruents: /b, p, t, d, k, g, f, v, θ, ð, s, z, ʃ, ʒ/. Five repetitions of each disyllable were included, for a total of 70 tokens per trial. For each subject, a different randomized list was used for each trial.

An audio-only pre-test was conducted prior to the main experiment to set subject-specific noise levels. Half of the subjects participated in the AT condition first, and half participated in the VT condition first, allowing for control of whether prior experience in one condition could affect performance in the other. A separate control trial was used to test participants' ability to perceive the relevant consonants based on audio-only or visual-only input (without tactile information). Thus, within AT and VT conditions, half of the subjects participated in the control trial first, while half participated in the test trial first.

(1) Pre-trial. Each subject wore isolating headphones playing white noise with a frequency band from 0 to 11 025 Hz at a sampling rate of 22 050 Hz, and was instructed to close his or her eyes. The subject sat at arm's length from the experimenter, who read from a randomized list of /aCá/ tokens prepared as described above. Simultaneously, a second experimenter adjusted the volume of the white noise in the headphones. Noise levels were set so that the subject could distinguish the correct segment with greater than chance frequency. During the pre-trial for the VT condition, each subject was instructed to raise a hand when he or she heard the experimenter speaking; noise levels were increased until the subject indicated (by a lowered hand) that the experimenter's voice was no longer audible at all. The final level was set slightly higher than this, and several additional utterances were used to confirm inaudibility.

(2) AT trial (With Tactile condition). Subjects wore isolating headphones playing white noise partially obscuring the acoustic speech signal as described above, and were instructed to keep their eyes closed for the duration of the trial (confirmed after the experiment based on the video record of each session). Each subject was seated at arm's length from the experimenter, with his or her right hand placed on the experimenter's face as per the Tadoma method: (a) the subject's index finger was placed horizontally just above the experimenter's mandibular ridge, (b) the other three fingers were fanned out beginning just below the mandibular ridge, across the experimenter's throat, (c) the palm was held over the experimenter's jaw and chin, and (d) the thumb was placed lightly on the experimenter's lips. Subjects received no instructions concerning how or whether to interpret the information conveyed through the hand-face contact. Once the subject was in place, the experimenter read aloud from a list of 70 token utterances prepared as described above. Despite many hours of training to ensure constant productions, the experimenter occasionally misspoke (as confirmed in the video and audio record of the experiment), with the result being that not all consonants in the stimuli lists were presented to each subject exactly five times. Subjects were instructed to listen to the experimenter, and to repeat each disyllable aloud in full to the best of his or her ability. A second experimenter recorded the subject's responses. This experimenter's record was later checked against the video and audio

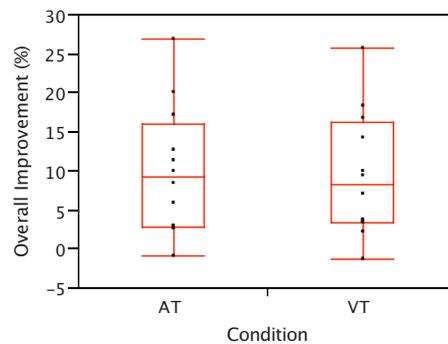


Fig. 1. (Color online) Overall percent improvement in correct segment identification due to tactile information in AT and VT conditions. Standard box plots show median, range and quartiles.

record by consensus among three phonetically trained experimenters. The audio record was also used to confirm that the experimenter spoke at a consistent amplitude across conditions.

(3) VT trial (With Tactile condition). Procedures were identical to the AT trial, except as follows: (a) the amplitude of the white noise was adjusted (as described above) to completely obscure the acoustic speech signal, (b) subjects kept their eyes open for the duration of the experiment, and (c) subjects were instructed to look at the primary experimenter throughout the trial.

(4) Control trials (Without Tactile conditions). Procedures were identical to the AT and VT trials, respectively, except that subjects had no physical contact with the experimenter during the trial.

3. Results

3.1 Overall results

Total percent improvement in consonant identification was compared across subjects and consonants for the Tactile conditions relative to the control conditions, i.e., auditory only or visual only, using paired t-tests (paired by subject). Results were highly significant, indicating a mean improvement of 9.96% (percentage points out of total possible correct, compared with a hypothesized difference of zero) for the AT condition [$t(11)=-4.2100$; $p=0.0015$] and 9.42% for the VT condition [$t(11)=-4.0868$; $p=0.0018$].

Total percentage improvement across VT and AT conditions was also compared using a paired t-test (paired by subject). Results indicate no difference between VT and AT conditions [$t(11)=0.1483$; $p=0.8848$]. Figure 1 plots improvements for all subjects with quartile box plots.

No significant effects of order of trial presentation were found, allowing all subjects to be included in all across-subject analyses.

3.2 Subject-by-subject results

Figure 2 shows percent accuracy results plotted by subject for AT and VT conditions. Although overall results were virtually identical across AT and VT conditions (Fig. 1), subject-specific results varied considerably. Figure 3 plots each subject's improvement in AT vs. VT conditions in a bivariate scatterplot. Linear regression indicates a significant negative correlation between improvement in AT vs. VT conditions ($r^2=0.4479$; $p=0.0243$; Subject 7's values lay outside of a 95% confidence ellipse, and were therefore omitted from this analysis). No significant correlation was found between baseline accuracy in a modality and improvement due to adding tactile information.

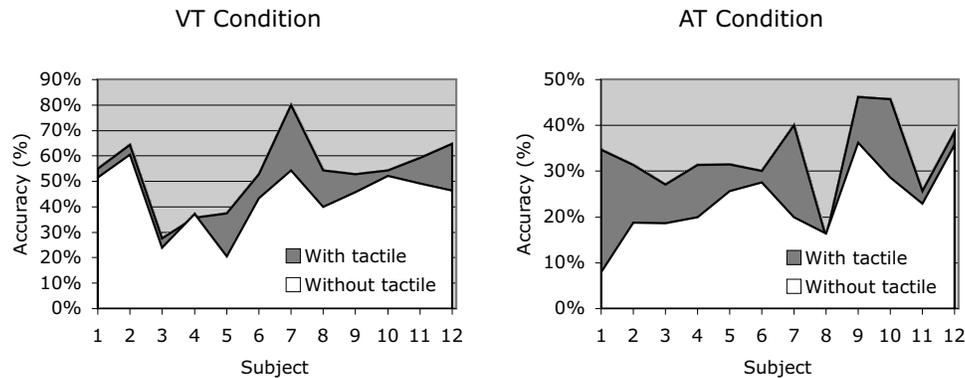


Fig. 2. Subject-by-subject improvement in correct segment identification in AT (left graph) and VT (right graph) conditions. The top line in each graph shows accuracy in With Tactile condition and the bottom line shows accuracy in Without Tactile condition. The gray region between the lines represents the improvement due to addition of tactile information.

4. Discussion

Across-subject results for the AT vs. VT conditions suggest that manual tactile information does enhance speech perception by about 10% in untrained perceivers. However, subject-specific results indicate that individual subjects vary substantially in how much improvement is gained from adding tactile information to each of the other perceptual modalities tested. Specifically, subjects who gain more from adding tactile information to the auditory modality tend to gain less from adding tactile information to the visual modality, and vice-versa. This supports the contention that modal additivity may differ for different cross-modal pairings across subjects.

5. Conclusion

The findings of this study support the notion that manual tactile information relevant to recovering speech gestures enhances speech perception in normal perceivers untrained in methods of tactile speech perception even for combinations of modalities where specific prior experience is unlikely (e.g., visual + tactile). Further, this enhancement occurs regardless of the other active modality—auditory or visual. However, the finding that subjects who gain more from tactile input coupled with an auditory baseline gain less from tactile input coupled with a visual baseline suggests that there is substantial individual variation in terms of which modality is favored for speech perception and how cross-modal information is combined.

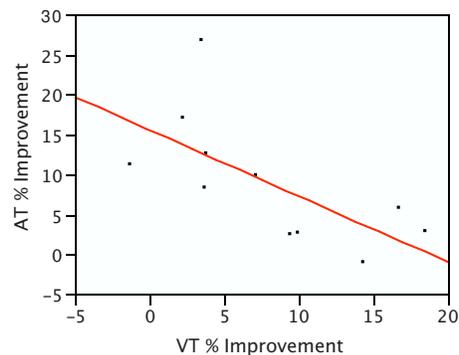


Fig. 3. (Color online) Relationship between percent improvement in correct segment identification due to adding tactile information. Linear fit shown indicates a significant negative correlation.

Acknowledgments

The authors thank Yoko Ikegami and Doug Whalen, and acknowledge support from an NSERC Discovery Grant to B.G. and NIH Grant No. DC-02717 to Haskins Laboratories.

References and links

- Alcorn, S. (1932). "The Tadoma method," *Volta Rev.* **34**, 195–198.
- Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (1989). "Speech perception using combinations of auditory, visual, and tactile information," *J. Rehabil. Res. Dev.* **26**(1), 15–24.
- Chomsky, C. (1986). "Analytic study of the Tadoma method: Language abilities of three deaf-blind subjects," *J. Speech Hear. Res.* **29**, 332–347.
- Fowler, C., and Dekle, D. (1991). "Listening with eye and hand: Crossmodal contributions to speech perception," *J. Exp. Psychol. Hum. Percept. Perform.* **17**, 816–828.
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature (London)* **264**, 746–748.
- McGrath, M., and Summerfield, Q. (1985). "Intermodal timing relations and audio-visual speech recognition by normal hearing adults," *J. Acoust. Soc. Am.* **77**, 676–685.
- Norton, S. J., Schultz, M. C., Reed, C. M., Braid, L. D., Durlach, N. I., Rabinowitz, W. M., and Chomsky, C. (1977). "Analytic study of the Tadoma method: Background and preliminary results," *J. Speech Hear. Res.* **20**, 574–595.
- Reed, C. M., Rubin, S. I., Braid, L. D., and Durlach, N. I. (1978). "Analytic study of the Tadoma method: Discrimination ability of untrained observers," *J. Speech Hear. Res.* **21**, 625–637.
- Reed, C. M., Doherty, M. J., Braid, L. D., and Durlach, N. I. (1982). "Analytic study of the Tadoma method: Further experiments with inexperienced observers," *J. Speech Hear. Res.* **25**, 216–223.
- Reed, C. M., Durlach, N. I., Braid, L. D., and Schultz, M. C. (1989). "Analytic study of the Tadoma method: Effects of hand position on segmental speech perception," *J. Speech Hear. Res.* **32**, 921–929.
- Sparks, D. W., Kuhl, P. K., Edmonds, A. E., and Gray, G. P. (1978). "Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of segmental features of speech," *J. Acoust. Soc. Am.* **63**(1), 246–257.