THE TEMPORAL WINDOW OF AUDIO-TACTILE INTEGRATION IN SPEECH PERCEPTION

Bryan Gick[1, 2] and Yoko Ikegami[1]
[1]Dept. of Linguistics, University of British Columbia, Vancouver, 2613 West Mall, British Columbia, Canada V6T1Z4
[2]Haskins Laboratories, 300 George St., 9th Fl., New Haven, CT 16511, USA

1. INTRODUCTION

Asynchronously presented auditory and visual information is integrated asymmetrically in speech perception [1, 2, 3]. For example, Munhall & al. [4] found that audio-visual integration of speech occurred even when the audio signal lagged the video signal by 240ms; however, when the audio signal preceded the video signal, perceivers only integrated 60ms of asynchrony. Munhall & al. suggest that this asymmetrical effect window may be attributable to perceivers' learned awareness of physical properties of the natural world (in this case, of the differing atmospheric speeds of sound and light): "This trend is not surprising since the relative speeds of sound and light would produce many natural occurrences of auditory events lagging their visual counterparts in the natural world" [4, p. 354]. However, this explanation has not been substantiated via comparison with other perceptual modalities.

Replication using the tactile modality should provide a test case for this question: Fowler & Dekle [5] and Gick *et al.* [6] found that untrained perceivers integrate tactile and auditory modalities through direct manual contact with speakers' faces. However, even if realistic and precisely timed synthetic facial (presumably robotic) stimuli could be constructed, this methodology would still fail to provide a natural signal transmission delay comparable to that of light or sound. The present experiment responds to this by coupling an acoustic speech signal with speech-like synthetic tactile stimuli in the form of small bursts of air following aspirated consonants.

The air speed of speech-like turbulent flow is considerably slower than that of sound in air, with flow velocity dropping off log-linearly after expulsion from the mouth [7]. If the physics-based hypothesis (i.e., the explanation based on perceivers' awareness of the relative physical transmission times of different signals) is correct, then the direction of asymmetry in the perceptual integration window should parallel the temporal difference between the relative speeds of sound and air flow. Any other result will fail to support the physics-based hypothesis.

2. METHOD

2.1 Participants

13 adult perceivers participated in the study. All were native speakers of English with no history of speech or hearing problems.

2.2 Stimuli

Acoustic stimuli consisted of recordings of 440 tokens of *pa* and *ba* produced in random order by a single female English speaker. Acoustic stimuli were output through the right channel of a Mac G4 sound card, mixed through a PreSonus mixing board with white noise (at a level such that subjects' baseline correct identification of *pa/ba* was at approximately 75%) and played to participants in stereo through Direct Sound Extreme Isolation headphones.

Tactile stimuli consisted of gentle bursts of air imparted via a vinyl tube at 7cm from the skin. Bursts were released from an air compressor at ~5psi using a Teknocraft 12-volt DC 2-way solenoid valve with a .032-inch orifice. The switch operating the solenoid valve was activated by a voltage initiated by an acoustic square wave output through the left channel a Mac G4 sound card amplified to 5 volts using a Frequency Devices voltage amplifier. Square waves were 60ms long (the average duration of aspiration for "pa" tokens used in the experiment), and offset leftward by 30ms to correct for a 30ms total system latency (see Figure 1).

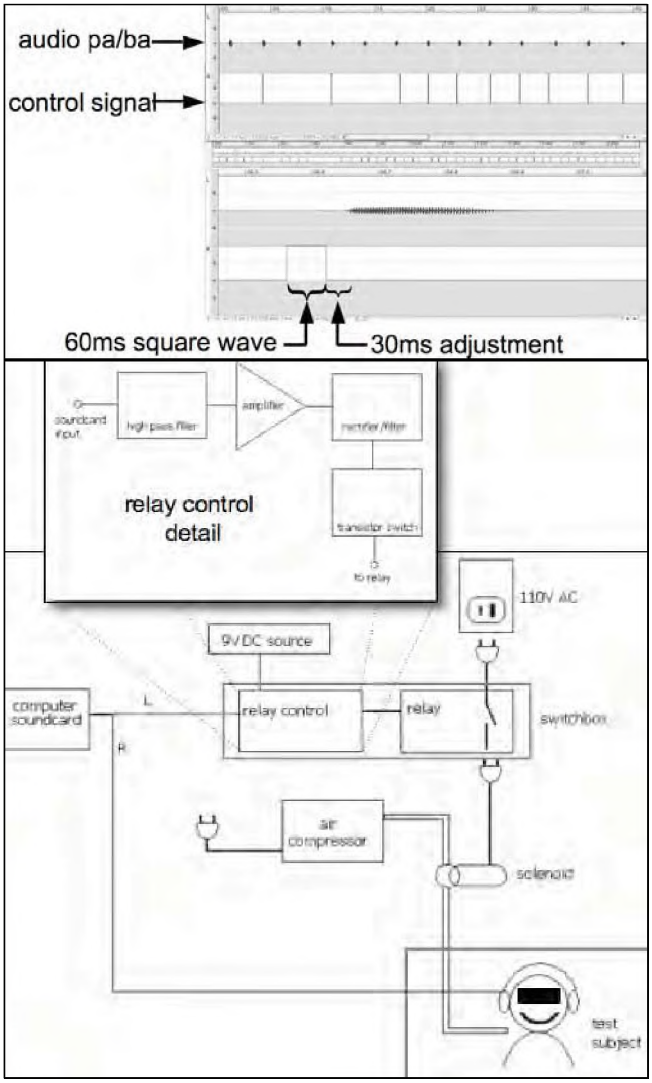

Fig. 1. Example of acoustic control and *pa/ba* signals (top) and flowchart of stimulus presentation system (bottom).

24 experimental conditions were tested, with temporal offsets between air bursts and spoken tokens varying by condition as follows: No Burst, 0ms (Simultaneous), ±50ms, ±100ms, ±200ms, ±300ms, and ±500ms (Distractor). Each

offsets occured in two different conditions (once with *pa* and once with *ba*). Participants heard 20 examples for each condition (except for the ±500ms distractors, for which there were only 10 items each), randomly distributed throughout the experiment, with one example per 3 seconds.

2.3 Procedure

Participants were seated in a soundbooth and read a script describing this experiment as testing their ability to identify different spoken syllables under conditions similar to those experienced by an airplane pilot. No specific mention was made of the air tube (indeed, some subjects reported not being aware of the air burst at all during the experiment). Participants were briefly trained to give forced-choice responses using a button box (with L/R responses balanced across participants), then blindfolded. Headphones were then placed on the participant, and the air tube put in place aiming at the right side of the neck.

3. RESULTS

Figure 2 shows the mean percent of correctly identified *pa* and *ba* syllables across subjects, plotted by condition. Paired t-tests (by subject) indicate significant enhancement to identification of *pa* responses with burst, and significant interference with identification of *ba* responses, in Simultanous conditions (compared with No Burst baseline conditions; $p > .05$). For both *pa* and *ba*, the effect at -50ms was not significantly different from Simultaneous; however, while the effect continued only to +50ms for *ba*, it persisted to a delay of +200ms for *pa* (as indicated by the circles in Figure 2).

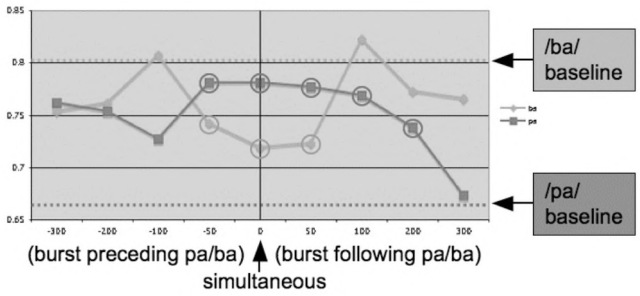

Fig. 2. Mean percent of correctly identified "pa" (dark grey line) and "ba" (light grey line) syllables. Circles indicate contiguous temporal offset conditions where percent identification did not differ significantly from Simultaneous.

4. DISCUSSION

In this experiment, tactile stimuli in the form of small bursts of air were directed at perceivers' necks while they heard productions of *pa* and *ba*. In baseline conditions, a burst occurring immediately prior to vowel onset (i.e., simultaneous with aspiration for *pa*) significantly enhanced perception of *pa* and significantly interfered with perception of *ba*. Asynchronous results showed a similar effect window to previous audio-visual studies: For asynchronously presented bursts, the temporal window of the enhancement effect of air bursts on perception of *pa* (but not the interference effect on *ba*) was asymmetrical, with integration occurring when the air burst followed the audio signal by 200ms, but only by 50ms when the air burst preceded the audio signal. The direction of this perceptual asymmetry parallels the temporal difference between the speeds of sound and air flow, supporting the physics-based hypothesis. Future work will attempt to address the question of whether perceivers' apparent understanding of physical properties of the world is learned or innate.

REFERENCES

[1] N. Dixon & L. Spitz. (1980). The detection of audiovisual desynchrony. *Perception,* 9, 719-721.

[2] P. M. T. Smeele, A. C. Sittig & V. J. Van Heuven. (1992). Intelligibility of audio-visually desynchronized speech: Asymmetrical effect of phoneme position. In *Proceedings of the International Conference on Spoken Language Processing*, pp. 65-68.

[3] Q. Summerfield. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London: Series B*, 335, 71-78.

[4] K. G. Munhall, P. Gribble, L. Sacco & M. Ward. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics,* 58(3), 351-362.

[5] C. A. Fowler, & D. J. Dekle. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816-828.

[6] B. Gick, K. Jóhannsdóttir, D. Gibraiel and J. Muehlbauer. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America,* 123(4), EL72-76.

[7] P. Anderson, D. Derrick, B. Gick & S. Green. (Under review). Characteristics of air puffs produced in English 'pa': Experiments and simulations. *Journal of the Acoustical Society of America.*

ACKNOWLEDGMENTS

This project benefited greatly from technical contributions of Gordon Ramsay and discussions with Douglas Whalen, and was funded by an NSERC Discovery Grant to Bryan Gick and NIH Grant DC-02717 to Haskins Laboratories.