Full body aero-tactile integration in speech perception

Donald Derrick[1], Bryan Gick [1,2]

[1]Department of Linguistics, University of British Columbia, Vancouver, BC V6T1Z4, Canada
[2]Haskins Laboratories, New Haven, Connecticut 06511-6695, USA
`dderrick@interchange.ubc.ca, gick@interchange.ubc.ca`

Abstract

We follow up on our research demonstrating that aero-tactile information can enhance or interfere with accurate auditory perception, even among uninformed and untrained perceivers [1]. Mimicking aspiration, we applied slight, inaudible air puffs on participants' skin at the ankle, simultaneously with syllables beginning with aspirated ('pa', 'ta') and unaspirated ('ba', 'da') stops, dividing the participants into two groups, those with hairy, and those with hairless ankles. Since hair follicle endings (mechanoreceptors) are used to detect air turbulence [2] we expected, and observed, that syllables heard simultaneously with cutaneous air puffs would be more likely to be heard as aspirated, but only among those with hairy ankles. These results demonstrate that information from any part of the body can be integrated in speech perception, but the stimuli must be unambiguously relatable to the speech event in order to be integrated into speech perception.

Index Terms: speech perception, aero-tactile integration, embodiment theory

1. Introduction

Visual information from a speaker's face has long been known to enhance [3] or interfere with [4] accurate auditory perception. This Auditory-visual integration has typically been attributed to the frequency and robustness with which perceivers jointly encounter event-specific information from these two modalities [5].

Previous studies on auditory-tactile integration had found an influence of tactile input on speech perception only under limited circumstances, either where perceivers were aware of the task [6], [7] or where they had received training to establish a cross-modal mapping [8, 9, 10]. Our recent study [1] demonstrated auditory-aerotactile integration even among uninformed and untrained perceivers, with data that is neither frequently nor robustly jointly encountered during speech except as air flow within the mouth of the speaker.

However, the results did not yet demonstrate that the entire body can be used to integrate auditory-aerotactile information in speech. To do so, we needed to demonstrate that the effect worked in a part of the body as far away from the ears as possible and as insensitive to tactile stimuli as possible. The side of the ankle was chosen because it represents the part of the body farthest away from the vocal tract that also has low tactile sensitivity [2]. Also, we deemed this to be one of the least likely locations where a perceiver might have prior experience feeling speech-related air puffs on the skin. As a result, we expected it to be the least likely part of the body to respond to aero-tactile stimuli.

For all the participants in our previous study, the neck and hand always had hair [1]. For cultural reasons, women in Vancouver tend to depilate their legs, so we also had the opportunity to test whether the absence of the ability to detect hair follicle motion would have an effect on auditory-aerotactile integration. Because low intensity turbulent airflow is generally detected as simply a cooling or pressure sensation on hairless skin, but as airflow on hairy skin, we expect that auditory-aerotactile integration will only work on those with hairy ankles. We therefore tested two groups, one with hairy ankles, and one with hairless ankles, under the assumption that proper perception of airflow requires hair follicles.

This experiment will help us identify whether the whole body can participate in the integration of information for speech perception, and how important it is that multi-modal information be unambiguously relatable to the same speech event. Because people sometimes speak to each other while standing close together, our previous study[1] left open the question of whether the integration was the result of specific previous experience with speech air flow information directed at the neck and hand. If we find that puffs on the ankle are integrated similarly to those on the hand or neck, this would support the position that this kind of cross-modal integration does not require previous location-specific experience.

2. Methods

The methods used to test this hypothesis closely match those for the experiments published in our previous research ([1]) as we used the same stimuli and setup, but with aero-tactile stimuli directed at the ankles instead of the neck or hand.

2.1. Participants

Both male and female subjects were tested. However, only male subjects had hairy ankles, and only female participants had hairless ankles. Our previous study showed no gender-specific response to the same stimulus at other body locations, so we're confident that any difference is strictly due to the presence/absence of body hair. Subjects were instructed to come with bare lower legs. We tested a total of 44 subjects, 22 subjects with hairless ankles, and 22 subjects with hairy ankles.

2.2. Stimuli and Apparatus

We created auditory stimuli by recording 8 repetitions of each of the syllables 'pa', 'ba', 'ta', and 'da' from a male native speaker of English, matching for duration (390-450 ms each), fundamental frequency (falling pitch from 90 Hz to 70 Hz), and intensity (normalized to 70 dBA). Subjects heard syllables in two separate blocks, one containing only labial consonants ('pa' and 'ba') and the other containing only alveolar consonants ('ta' and
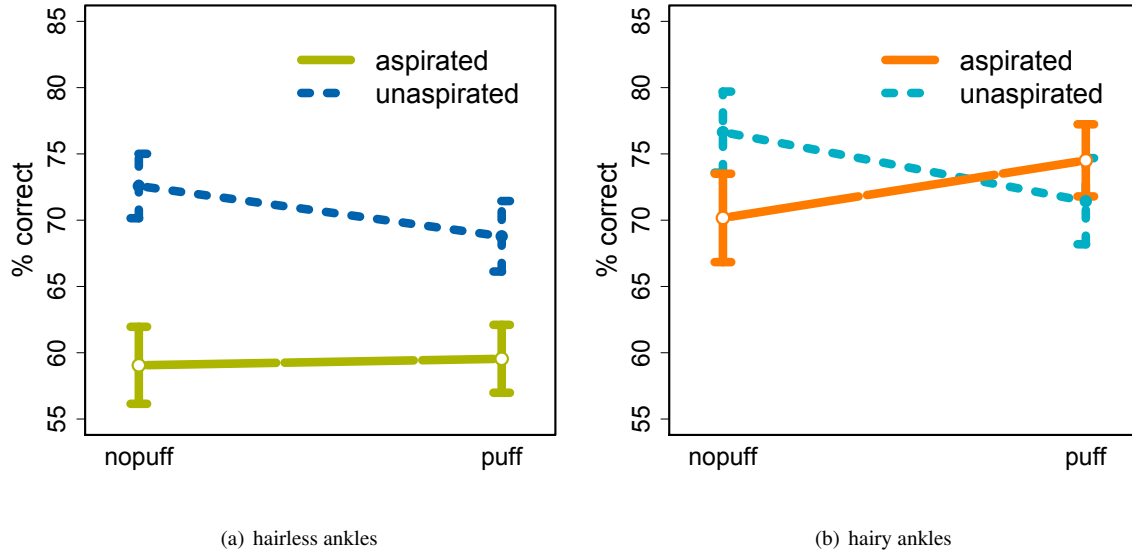
26 – 30 September 2010, Makuhari, Chiba, Japan

| | (a) hairless ankles | | (b) hairy ankles |

Figure 1: 2x2 Interaction graphs with standard error bars. Dependent variable group 1: Aspirated = auditory 'pa' or 'ta', Unaspirated = auditory 'ba' or 'da'. Dependent variable group 2: puff vs. no puff (on the ankle). Independent variable: % correct = % answers matching auditory stimuli.

'da'). The 16 unique tokens in each block were heard 4 times each, twice as auditory-only controls and twice paired with tactile stimuli. Auditory stimuli were accompanied by white noise played at a volume intended to reduce the overall accuracy of token identification and so generate significant ambiguity; actual accuracy is documented in Tables 1 and 2.

We used a solenoid valve attached to an air compressor to synthesize small puffs of air designed to replicate the pressure profile (transient boundary condition), high frequency noise, low frequency 'pop' duration and temporal relation to vowel onset of natural speech aspiration. The airflow device used to produce the synthetic puffs consisted of a 3-gallon Jobmate oil-less air compressor connected to an IQ Valves on-off 2-way solenoid valve (model W2-NC-L8PN-S078-MB-W6.0-V110) connected to a Campbell Hausfeld MP513810 air filter, which reduced the sound volume conducted through the 1/4-inch vinyl tubing. The tubing was passed through a cable port into the soundproof room and mounted on a microphone boom-stand. The synthetic puff airflow was quickly turbulent upon leaving the tube, with an average turbulence duration of 84 ms, compared to 60 ms voice onset time (VOT) for our speakers average (mean) 'pa', and close to the VOT range of 54-80 ms for English word-onset voiceless (aspirated) stops[11].

The output pressure of the synthesized puffs was adjusted so that impact was minimally perceptible by subjects. Nevertheless, the aero-tactile stimuli was slightly stronger than the original study, as the pressure was higher (9 psi [632.76 g cm^{-2}] instead of 6 psi [421.84 g cm^{-2}]), and the source closer to the skin (5 cm instead of 8 cm).

In both experiments, air puffs were applied cutaneously on the distal surface of the ankle just above the lateral malleolus.

A single stereo audio signal supplied both the auditory stimuli heard by subjects and the activation signal to open the air valve. The right channel carried the spoken syllables to both ears through headphones worn by subjects, while the left channel activated the solenoid by outputting 50ms-long 10-kHz sine waves at the maximum amplitude of the computers sound card (1V) through a voltage amplifier to a relay. The sine waves were time-aligned with the speech signal such that, after correction for system latency, air puffs exited the tube starting 50 ms prior to vowel onset and ending at the moment of vowel onset, simulating the timing of naturally produced English aspirated consonants.

For stimuli presentation, a custom-built computer program written in Java 1.6 recorded responses from a customized keypad and presented new tokens 1500 ms after each response.

2.3. Procedure

Prior to the experiment, subjects were told that they might experience background noise and unexpected puffs of air. Subjects who did not come with bare lower legs were asked to roll up pants or remove boots if necessary, and they were seated in a soundproof booth. Subjects were then blindfolded and provided with auditory stimuli through sound-isolating headphones. Setup of equipment to deliver tactile stimuli was completed after the subjects were blindfolded to conceal the body location of air puffs. Once positioned and ready, subjects were asked to identify by button press whether they heard 'pa' or 'ba' in the labial block; 'ta' or 'da' in the alveolar block.

Half the participants received the labial ('pa', 'ba') block first, and half received the alveolar ('ta', 'da') block first. Within each block, participants heard 12 randomized practice tokens (6 with and 6 without air puffs, (no feedback provided) followed by 16 randomized experimental tokens for each condition (aspirated vs. unaspirated, puff vs. no puff, randomized), totalling 64 experimental tokens per block. Half the subjects were asked to press the left button of our customized keypad to indicate an aspirated response, and half the right button.

3. Results

The results of the experiments on those with hairless ankles (experiment 1) and hairy ankles (experiment 2) are presented below:

3.1. Experiment 1: hairless ankles

A 2x2 repeated measures ANOVA with aspiration (aspirated vs. unaspirated) and air puffs (present vs. absent) as factors were conducted on the combined data from the alveolar and labial blocks of the hairless ankle study. Descriptive statistics are found in Table 1.

Table 1: *Descriptive statistics for the hairless ankle experiment*

Condition	Mean	Std. Deviation	N
unaspirated, no puff	72.6%	11.4%	22
unaspirated, puff	68.8%	12.5%	22
aspirated, no puff	59.1%	13.6%	22
aspirated, puff	59.5%	12.0%	22

The results showed that the interaction of air puffs with the perception of aspiration was not significant (α = .05, $F_{(1,21)}$ = 1.1631, p = .29, partial eta-squared (Pt η^2) = 5.2%). The interaction graph is shown in Fig. 1(a).

3.2. Experiment 2: hairy ankles

A 2x2 repeated measures ANOVA with aspiration (aspirated vs. unaspirated) and air puffs (present vs. absent) as factors were conducted on the combined data from the alveolar and labial blocks of the hairy ankle study. Descriptive statistics are found in Table 2.

Table 2: *Descriptive statistics for the hairy ankle experiment*

Condition	Mean	Std. Deviation	N
unaspirated, no puff	76.6%	14.3%	22
unaspirated, puff	71.4%	15.2%	22
aspirated, no puff	70.2%	15.6%	22
aspirated, puff	74.5%	12.7%	22

The results showed that the interaction of air puffs with the perception of aspiration was significant (α = .05, $F_{(1,21)}$ = 6.3521, p = .020, Pt η^2 = 23.2%). The interaction graph is shown in Fig. 1(b).

3.3. 2x2x2 mixed ANOVA

A 2x2x2 mixed ANOVA comparing the interaction of aspiration, presence or absence of air puffs to the ankle, and presence or absence of hair on the ankles shows that in this experiment, participants with hair on the ankle were more likely to perceive all stimuli accurately than those without (Error: subject, $F_{(1,38)}$ = 8.84, p = .005, Pt η^2 = 19.7%). Unaspirated stimuli was perceived more accurately than aspirated stimuli (Error: subject * aspiration, $F_{(1,38)}$ = 6.93, p = .012, Pt η^2 = 15.4%). The interaction between voicing and air puff was significant (Error: subject * aspiration * puff, $F_{(1,42)}$ = 6.22, p = .017, Pt η^2 = 12.9%). However, three-way interaction was not significant.

4. Discussion

The results show that air puffs on the ankles influence speech perception in a similar way to that documented in our previous paper [1]. Participants with hairy ankles perceived the stimuli more accurately than those without. The results also show that unaspirated stimuli was perceived more accurately than aspirated stimuli.

Direct comparison of the hairy and hairless ankle results shows significant speech perception influence only among participants with hairy ankles. Hairless ankles limited the ability of the participants to enjoy the advantage of aero-tactile stimuli, demonstrating the need for hair follicle displacement, and by inference the ability to detect the air flow as turbulence and not just pressure or temperature change. That is, auditory-aerotactile integration is a full-body process, but requires event-relevant information in order to fully affect auditory perception.

These results show that while perceivers need little or no prior location-specific experience with the joint stimuli, the stimuli must be event-specific in order to enhance or interfere with speech perception. That is, the stimuli must be grounded in the physics of speech, and perceivable as such.

5. Acknowledgements

6. References

[1] B. Gick and D. Derrick, "Aero-tactile integration in speech perception," Nature, vol. 462, pp. 502–504, 26 November 2009, doi:10.1038/nature08572.

[2] S. Weinstein, The Skin Senses. Springfield, Illinois: Charles C Thomas, 1968, ch. Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality, pp. 195–222.

[3] W. H. Sunby and I. Pollack, "Visual Contribution to Speech Intelligibility in Noise," Journal of the Acoustical Society of America, vol. 26, pp. 212–215, 1954.

[4] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," Nature, vol. 264, pp. 746–748, 1976.

[5] R. L. Diehl and K. R. Kluender, "On the objects of speech perception," Ecological Psychology, vol. 1, no. 2, pp. 121–144, 1989.

[6] C. A. Fowler and D. J. Dekle, "Listening with eye and hand: Crossmodal controbutions to speech perception," Journal of Experimental Psychology, Human Perception and Performance, vol. 17, pp. 816–828, 1991.

[7] B. Gick, K. M. Jóhannsdóttir, D. Gibraiel, and J. Mühlbauer, "Tactile enhancement of auditory and visual speech perception in untrained perceivers," Journal of the Acoustical Society of America, vol. 123, no. 4, pp. EL72–EL76, 2008.

[8] D. W. Sparks, P. K. Kuhl, A. E. Edmonds, and G. P. Gray, "Investigating the MESA (Multipoint Electrotactile Speech Aid): the transmission of segmental features of speech," Journal of the Acoustical Society of America, vol. 64, p. 246, 1978.

[9] C. M. Reed, N. I. Durlach, L. D. Braida, and M. C. Schultz, "Analytic study of the Tadoma method: Effects of hand position on segmental speech perception," Journal of Speech, Language and Hearing Research, vol. 32, no. 4, p. 921, 1989.

[10] L. E. Bernstein, M. E. Demorest, D. C. Coulter, and M. P. O'Connell, "Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects," Journal of the Acoustical Society of America, vol. 90, no. 6, pp. 2971–2984, 1991.

[11] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," Word, vol. 20, pp. 384–422, 1964.