Published in final edited form as:

Multisens Res. 2013 ; 26(5): 405–416.

Aerotactile Integration from Distal Skin Stimuli

Donald Derrick^{1,2,*} and Bryan Gick^{3,4,**}

¹University of Western Sydney, MARCS Institute, Locked Bag 1791, Penrith, NSW, 2751, Australia ²University of Canterbury, New Zealand Institute of Language, Brain & Behaviour, Private Bag 4800, Christchurch 8140, New Zealand ³Department of Linguistics, University of British Columbia, Totem Field Studios, 2613 West Mall, Vancouver, BC, Canada, V6T 1Z4 ⁴Haskins Laboratories, New Haven, Connecticut, 06511, USA

Abstract

Tactile sensations at extreme distal body locations can integrate with auditory information to alter speech perception among uninformed and untrained listeners. Inaudible air puffs were applied to participants' ankles, simultaneously with audible syllables having aspirated and unaspirated stop onsets. Syllables heard simultaneously with air puffs were more likely to be heard as aspirated. These results demonstrate that event-appropriate information from distal parts of the body integrates in speech perception, even without frequent or robust location-specific experience. In addition, overall performance was significantly better for those with hair on their ankles, which suggests that the presence of hair may help establish signal relevance, and so aid in multi-modal speech perception.

Keywords

Multi-modal speech perception; auditory perception; tactile perception

1. Introduction

Visual information from a speaker's face has long been known to enhance (Sumby and Pollack, 1954) or interfere with (McGurk and MacDonald, 1976) 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 (Diehl and Kluender, 1989).

Previous studies on auditory-tactile integration had found an influence of tactile input on speech perception. Initially these results were under limited circumstances, either where perceivers were aware of the task (Fowler and Dekle, 1991; Gick *et al.*, 2008) or where they had received training to establish a cross-modal mapping (Bernstein *et al.*, 1991; Reed *et al.*, 1989; Sparks *et al.*, 1978). A later study (Gick and Derrick, 2009) showed that uninformed and untrained perceivers were able to integrate auditory speech information with airflow

*To whom correspondence should be addressed. donald.derrick@gmail.com.
**gick@mail.ubc.ca

directed at one of two body locations: the neck or hand (locations where perceivers are likely to have had at least some prior experience throughout their lives feeling airflow from their own or others' speech). However, previous studies have not tested whether puffs applied at more remote body locations could be perceived as aerotactile information in speech, leaving open the possibility that location-specific experience is a necessary component of auditory–aerotactile integration.

The present study tests whether integration is possible *via* puffs delivered to a part of the body that is (1) maximally distant from the ears, both physically and within the somatosensory cortex (see Penfield and Rasmussen, 1950), and (2) where perceivers have had minimal previous experience feeling speech-related airflow. The part of the body that best fit these criteria was the outside of the ankle. The ankle is extremely unlikely to have previously received robust speech airflow, except possibly sporadically during infancy; partly because the ankle is at the opposite end of the body from the speech apparatus, and partly because the ankle has such low tactile sensitivity (Weinstein, 1968), any speech airflow information that might have reached the ankles over the perceiver's lifetime has the least chance of being perceived as such.

Previous research also found that, while air puffs provide tactile information that integrates into speech perception, taps on the skin do not (see supplemental material Gick and Derrick, 2009). In addition to controlling for attention effects, this result suggests that the tactile signal has to be perceived as event-relevant, as opposed to merely synchronous (see Gick *et al.*, 2010, for the temporal window of aerotactile integration) in order to integrate with the auditory signal.

This is particularly important for the ankle because during infancy, which is the most likely time of auditory co-experience, hair growth on the ankles is extremely limited, and this may reduce the perception of airflow. Airflow on the skin and hair can generate many distinct percepts: Evidence shows airflow stimuli can generate sensations of cold (Todde, 2000), hair displacement sensory reactions (Hamalainen *et al.*, 1985), and responses in the remaining four mechanoreceptor types (Mizobuchi *et al.*, 2000). Hair and hair mechanoreceptors are important for detecting airflow direction properly, especially at low airflow pressures. Hair is so useful in helping identify airflow direction that bats use it for flight (Sterbing-D'Angelo *et al.*, 2011).

Without the hair displacement sense, it is generally believed airflow is more likely to be detected as pressure or coldness alone, and not airflow (see Todde, 2000). Hairy skin also has sensors devoted to pleasant touch, and the influence of these on airflow sensation is unknown and as a result potentially relevant (Löken, 2009). While it is unknown if these sensory differences matter, hairy skin is known to detect low-pressure airflow more readily than glabrous skin (Hamalainen *et al.*, 1985). We therefore also tested whether the presence or absence of ankle hair significantly affected auditory–aerotactile integration in speech.

For all the participants in our previous studies, the neck (Gick and Derrick, 2009; Gick *et al.*, 2010) and hand (Gick and Derrick, 2009) always had hair. In contrast, many North

Americans depilate their legs, giving us the opportunity to test whether the absence of the ability to detect hair follicle motion would have an effect on auditory–aerotactile integration.

The main hypothesis of this experiment is that erotactile integration in speech perception will occur with air puffs to the ankle. The secondary hypothesis is that hair on the ankles will aid in erotactile integration in speech perception, and may be necessary for establishing signal relevance.

2. Methods

The methods used in the present study were designed to replicate those of Gick and Derrick (2009), but with erotactile stimuli directed at the ankle instead of the neck or hand.

Note that we did not use naturally glabrous skin as a control for two reasons: The first is that glabrous skin is simply too sensitive to be suitable as a comparison site to the ankle. The second and most important reason is that glabrous skin is missing the a-beta hair follicle receptors that respond to hair displacement. We wanted to make the comparison between the two groups as similar as possible in order to avoid confounds.

It was also important to make sure participants could not hear the activation of the air puff device. Previous research using the same apparatus demonstrated that participants could not hear the activation of the air puffs even when discharged immediately adjacent to the headphones. For the present paper, we follow the same methods used in our original erotactile integration paper (Gick and Derrick, 2009), but with slight adjustments for application at the ankle location. The present design used a slightly higher air pressure (9 psi vs. 6 psi) and shorter distance to the skin surface (5 cm vs. 8 cm) than previous studies as pilot data suggested the original airflow could not be perceived at the ankle at all. While the slight increase in air pressure negligibly increases the volume of sound from our air puff device, the extreme distal body location placed the source of the puff an order of magnitude farther from the participant's head. For all experiments, we used the same DirectSound Extreme Isolation Ex-29 sound-isolating headphones, which have a United States Environmental Protection Agency noise reduction rating (NRR) of 29 dB.

2.1. Participants

Research participants for this study were drawn from the student population at the University of British Columbia, and were primarily undergraduates. Participants were asked about their hearing, and we used only those that reported having normal hearing. Participants were categorized into two groups, one *With Hair*, and the other *Without Hair* on their ankles, based on visual inspection of the ankle and leg. While not explicitly planned this way, all of the people with hair were male and all the people without hair were female. We tested a total of 44 participants, 22 participants with hair, and 22 participants without hair.

The current statistical analysis included the data from previous erotactile stimuli directed at the neck and hand (Gick and Derrick, 2009), which adds 22 participants for each experiment, and provides a total of 88 participants for the analysis. This data allows

comparison between areas of aerotactile stimulation, and makes the statistical analysis more meaningful and robust against errors.

2.2. Stimuli and Apparatus

We created auditory stimuli by recording eight repetitions of each of the syllables ‘pa’, ‘ba’, ‘ta’, and ‘da’ from a male native speaker of English, carefully controlling for duration (390–450 ms each), fundamental frequency (falling pitch from 90 Hz to 70 Hz), and intensity (normalized to 70 dBA). Perceivers heard syllables in two separate blocks, one containing only labial consonants (‘pa’ and ‘ba’) and the other containing only alveolar consonants (‘ta’ and ‘da’). The ‘ba’ and ‘da’ tokens form the *Unaspirated* sounds, and the ‘pa’ and ‘ta’ tokens form the *Aspirated* sounds. The 16 unique tokens in each block (total 32 unique sounds) were heard four times each, twice as auditory-only controls and twice paired with tactile stimuli in the form of air puffs on the skin, forming the *Puff vs. No Puff* conditions. Auditory stimuli were accompanied by white noise played at a volume intended to reduce the overall accuracy of token identification and so to generate significant ambiguity. Because the overall ambiguity from Gick and Derrick (2009) of 73.2% was suitable, we used the same SNR for this research.

We used a solenoid valve attached to an air compressor to synthesize small puffs of air designed to simulate 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 two-way solenoid valve (model W2-NC-L8PN-S078-MB-W6.0-V110).

Puffs were delivered to the participant *via* 1/4-inch vinyl tubing after passing through a Campbell Hausfeld MP513810 air filter, which reduced sound volume from the puff. The tubing was passed through a cable port into the soundproof room and mounted on a microphone boom-stand.

Airflow pressure was 9 psi [632.76 g cm^{-2}], presented from the 1/4-inch vinyl tube 5 cm away from the leg. As a result, the area of initial impact had a diameter of 2–3 cm (Derrick *et al.*, 2009). Air puffs were applied cutaneously on the distal surface of the ankle just above the lateral malleolus, as seen in Fig. 1.

A single stereo audio signal supplied both the auditory stimuli heard by participants and the activation signal to open the air valve. The right channel carried the spoken syllables to both ears through headphones worn by participants, while the left channel activated the solenoid by outputting 50-ms long 10-kHz sine waves at the maximum amplitude of the computer's sound card (1 V) 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. 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 speaker's average (mean) ‘pa’, and close to the VOT range of 54–80 ms for English word-onset aspirated stops (Lisker and

Abramson, 1964). Note that the airflow control is 50 ms long at the source, but an average of 84 ms long at the point of skin contact due to time-lag in pressure reduction within the system, acting much like the pressure decrease that exists in natural speech (Derrick *et al.*, 2009).

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, participants were told that they might experience conditions familiar to small aircraft pilots, including background noise and unexpected puffs of air. Participants who did not come with bare lower legs were asked to roll up pant legs or remove boots if necessary, and they were seated in a soundproof booth with their right foot resting on a wooden pedal, which stabilized their ankle position. Participants were then blindfolded and provided with auditory stimuli through a pair of sound-isolating headphones. Setup of equipment to deliver tactile stimuli was completed after the participants were blindfolded to conceal the body location of air puffs. Participants were then asked to identify by button press whether they heard 'pa' or 'ba' in the labial block; 'ta' or 'da' in the alveolar block.

Blocks were counterbalanced: 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 (six with and six without air puffs, no feedback provided) followed by 16 randomized experimental tokens for each condition (Aspirated *vs.* Unaspirated, Puff *vs.* No Puff, randomized), totaling 64 experimental tokens per block. Half the participants were asked to press the left button of our customized keypad to indicate an aspirated response, and half the right button.

3. Data Analysis and Results

The results show that participants correctly identified tokens, based on the underlying acoustic stimuli, about 69.1% of the time. Full details can be seen in Table 1.

Generalized linear mixed-effects models (GLMM) were run on the interaction between aspiration [aspirated ('pa' and 'ta') *vs.* unaspirated ('ba' and 'da') stops], artificial air puff (present *vs.* absent), place of articulation [labial ('pa' and 'ba') *vs.* alveolar ('pa' and 'ba') stops], and puff location, including the hand, neck data from Gick and Derrick (2009), and the ankle data from this study, and gender (which also captures the distinction between ankles with and without hair).

Model fitting was then performed in a stepwise backwards iterative fashion, and models were back-fit along the Bayesian information criterion (BIC), to measure quality of fit. This technique allows for the isolation of a statistical model that provides the best fit for the data, and allows elimination of interactions in a statistically appropriate manner. The final model was:

$$Response \sim Location + Gender + (Aspiration * Puff) + (1 + Aspiration | Participant) + (1 | Sound) .$$

In this model, the response, as measured by accuracy in relation to the underlying auditory signal, was compared to the fixed effects. These fixed effects include (1) the location of the air puff (hand, neck, ankle), (2) the gender (male or female), (3) the interaction of aspiration in the acoustical signal and air puff. The model also factors in random effects, including (1) participant and speech aspiration, and (2) the underlying auditory signal. The results can be seen in Table 2.

The results show that participants in the ankle condition were less accurate than participants in the neck condition. Re-leveling the condition allowed the comparison of participants in the neck condition compared to participants in the hand condition, which were not found to be significantly different. The results also demonstrate male participants were more accurate than female participants, largely due to the higher accuracy of ankle participants with hair compared to those without hair.

Unaspirated tokens were perceived more accurately than aspirated tokens. Tokens with synchronous air puffs were perceived more accurately than tokens without synchronous air puffs.

Figure 2 shows the interaction between aspiration and the application of mechanical air puffs. Figure 2A, like the analytical statistics above, groups neck, hand, and ankle results because there were no significant three- or four-way interactions involving location or gender. The interaction results in Fig. 2A show that without the air puff, perceivers are much more accurate at identifying unaspirated tokens, and much less accurate at identifying aspirated tokens. However, when given an air puff, accuracy does not differ between aspirated and unaspirated tokens. In addition, Fig. 2B shows a similar pattern for the combined ankle data. Figure 2C and 2D show the interaction plots for the Hair and Without Hair groups. Note that despite the appearance of these graphs, stepwise model fitting showed no significant difference in the interaction patterns for the two ankle groups.

Random effects were used to account for individual participant and item variation. The random effects can be seen in Table 3.

4. Discussion

The results show that air puffs on the ankle influences speech perception in a similar way to those applied to the neck (Gick and Derrick, 2009; Gick *et al.*, 2010) and hand (Gick and Derrick, 2009). These interaction results support the argument that air puffs act as a source of speech information along with the acoustic signal, causing participants to be more likely to perceive aspiration in the initial consonant of both acoustically aspirated and unaspirated tokens. This, along with the observation that tokens with mechanical air puffs were perceived more accurately than those without, removes the possibility that air puffs act as a distractor that simply negatively influences speech perception. In addition, unaspirated tokens were perceived more accurately than aspirated tokens, and males perceived tokens

more accurately than females — largely due to a higher overall degree of accuracy in this ankle experiment.

While the descriptive statistics, and several intermediate models in the stepwise iterative process, suggested that integration was more effective for those with hair on their ankles than those without, these results did not produce the best fit for the data. The analytical tests simply show that aerotactile integration occurred with air puffs at any part of the body, be it the hands, neck, or ankles, and that male participants — those with hair on their ankles — were more accurate overall.

Nevertheless, an examination of the literature suggests there is much to learn about how people perceive airflow of various pressures, volumes, and velocities on normally hairy skin both with and without the hair removed. In addition, while it is known that airflow elicits reactions from all four mechanoreceptors, and specifically from hair mechanoreceptors, as well as cold sensors, little is known about how perceivers describe those sensations. Whether hair itself enhances aerotactile integration in speech perception is a question that maybe solved with future research where the same participants are tested before and after hair removal in order to avoid the potentially confounding interaction between gender and aerotactile integration.

The question of exactly how the brain integrates aerotactile information with other modalities in speech perception remains wide open for research. As with any area of basic human perception research, uncovering the mechanisms and extent of aerotactile integration in speech perception will create opportunities for novel medical devices and commercial products. We are currently working on these and related research questions.

At this point, we have confirmed that auditory–aerotactile integration is a full-body process in which event-relevant information presented synchronously influences auditory perception. These results demonstrate that information from distal parts of the body integrates in speech perception, even though listeners do not have frequent or robust experience with speech airflow directed towards their ankles. That is, prior location-specific aerotactile experience is not necessary for aerotactile integration in speech perception.

Acknowledgments

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Council of Canada (NSERC) to the second author, and by National Institutes of Health (NIH) Grant DC-02717 to Haskins Laboratories. The authors gratefully acknowledge the contributions of Laurie McLeod for help setting up the UBC system, Sylvia Renardy for organizing protocols, D. H. Whalen for helpful discussion, and students of the UBC ISRL for running participants and other assistance. We also want to thank our infant research colleagues for pointing out potential infant experience with aerotactile stimulation at the ankle. Special thanks to Patrick LaShell for help with our statistical analysis during the review process — his help made this a much better paper.

References

- Bernstein LE, Demorest ME, Coulter DC, O'Connell MP. Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *J Acoust Soc Am.* 1991; 90:2971–2984. [PubMed: 1838561]
- Derrick D, Anderson P, Gick B, Green S. Characteristics of air puffs produced in English 'pa': Experiments and simulations. *J Acoust Soc Am.* 2009; 125:2272–2281. [PubMed: 19354402]

- Diehl RL, Kluender KR. On the objects of speech perception. *Ecol Psychol.* 1989; 1:121–144.
- Fowler CA, Dekle DJ. Listening with eye and hand: Crossmodal contributions to speech perception. *J Exp Psychol -Human.* 1991; 17:816–828.
- Gick B, Derrick D. Aero-tactile integration in speech perception. *Nature.* 2009; 462:502–504. [PubMed: 19940925]
- Gick B, Ikegami Y, Derrick D. The temporal window of audio-tactile integration in speech perception. *J Acoust Soc Am -E L.* 2010; 128:EL342–EL346.
- Gick B, Jóhannsdóttir KM, Gibrael D, Mühlbauer J. Tactile enhancement of auditory and visual speech perception in untrained perceivers. *J Acoust Soc Am.* 2008; 123:EL72–EL76. [PubMed: 18396924]
- Hamalainen HA, Warren S, Gardner EP. Differential sensitivity to airpuffs on human hairy and glabrous skin. *Somatosens Res.* 1985; 2:281–302. [PubMed: 4035149]
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. *Word.* 1964; 20:384–422.
- Löken, LS. PhD Thesis. Sahlgrenska Academy, University of Gothenburg; 2009. A pathway for pleasant touch: linking peripheral receptors to central processing and hedonic experience.
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature.* 1976; 264:746–748. [PubMed: 1012311]
- Mizobuchi K, Kuwabara S, Toma S, Nakajima Y, Ogawara K, Hattori T. Single unit responses of human cutaneous mechanoreceptors to air-puff stimulation. *Clin Neurophysiol.* 2000; 111:1577–1581. [PubMed: 10964067]
- Penfield, W.; Rasmussen, T. *The Cerebral Cortex of Man.* Macmillan; New York NY, USA: 1950.
- Reed CM, Durlach NI, Braida LD, Schultz MC. Analytic study of the Tadoma method: Effects of hand position on segmental speech perception. *J Speech Lang Hear R.* 1989; 32:921.
- Sparks DW, Kuhl PK, Edmonds AE, Gray GP. Investigating the MESA (multipoint electrotactile speech aid): The transmission of segmental features of speech. *J Acoust Soc Am.* 1978; 64:246. [PubMed: 632417]
- Sterbing-D'Angelo S, Chadha M, Chiu C, Falk B, Xian W, Barcelo J, Zook JM, Moss CF. Bat wing sensors support flight control. *Proc Natl Acad Sci USA.* 2011; 108:11291–11296. [PubMed: 21690408]
- Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *J Acoust Soc Am.* 1954; 26:212–215.
- Todde V. Perception and sensitivity to horizontal turbulent air flows at the head region. *Indoor Air.* 2000; 10:297–305. [PubMed: 11089333]
- Weinstein, S. Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. In: Kenshalo, DR., editor. *The Skin Senses.* Charles C. Thomas; Springfield, IL, USA: 1968. p. 195-222.

Figure 1.
Location of air puffs on the distal surface of the ankle just above the lateral malleolus. This figure is published in colour in the online version.

Figure 2. Interaction of aspiration and mechanical air puff. % Accuracy = % answers matching auditory stimuli. (A) All data combined; (B) ankle combined; (C) Ankle-hair; (D) Ankle-without hair. This figure is published in colour in the online version.

Table 1

Descriptive Statistics: Aspiration = unaspirated ('ba' and 'da') vs. aspirated ('pa' and 'ta'). Air puff = artificial air puff vs. no air puff. Place = labial ('ba' and 'pa') vs. alveolar ('da' and 'ta'). Hair = hair on ankle vs. no hair on ankle. Accuracy = % of tokens accurately identified based on underlying acoustic signal. Std. err. = Standard error

Hair	Place	Aspiration	Puff	Accuracy	Std. err.
Hair	Alveolar	Aspirated	Puff	77.8%	3.60%
			No puff	72.2%	4.43%
		Unaspirated	Puff	75.1%	3.74%
			No puff	78.2%	4.28%
	Labial	Aspirated	Puff	70.9%	3.48%
			No puff	68.2%	3.73%
		Unaspirated	Puff	75.0%	3.66%
			No puff	75.0%	3.66%
Without hair	Alveolar	Aspirated	Puff	62.0%	3.86%
			No puff	61.0%	4.35%
		Unaspirated	Puff	68.0%	3.46%
			No puff	73.4%	3.12%
	Labial	Aspirated	Puff	57.2%	2.99%
			No puff	57.2%	2.70%
		Unaspirated	Puff	69.5%	2.55%
			No puff	71.6%	3.13%

Table 2

Fixed effects estimates based on the final model: $Response \sim Location + Gender + (Aspiration * Puff) + (1 + Aspiration|Participant) + (1|Sound)$. Table shows standard deviation of participant intercept, aspiration by participant, and underlying token intercept. 11 160 observations, 88 participants, and 32 sounds

	Estimate	Std. err.	z-value	P-value (> z)
(Intercept)	0.47	0.14	3.32	<0.001
Location — hand	0.21	0.13	1.55	0.121
Location — neck	0.32	0.14	2.35	*0.019
Gender — male	0.23	0.11	2.02	*0.044
Aspiration — unaspirated	0.67	0.18	3.74	*<0.001
Puff — puff	0.33	0.06	5.39	*<0.001
Aspirated * puff	0.69	0.09	7.82	*<0.001

Table 3

Random effects estimates based on the final model: $Response \sim Location + Gender + (Aspiration * Puff) + (1 + Aspiration|Participant) + (1|Sound)$. Table shows standard deviation of participant intercept, aspiration by participant, and underlying token intercept. 11 160 observations, 88 participants, and 32 sounds

	Variance	Std. dev.
Participant (intercept)	0.383	0.619
Aspiration — unaspirated	0.936	0.967
Sound (intercept)	0.134	0.367