Published in final edited form as:
Nature. 2009 November 26; 462(7272): 502–504. doi:10.1038/nature08572.

Aero-tactile integration in speech perception

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

[1]Department of Linguistics, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

[2]Haskins Laboratories, New Haven, Connecticut 06511-6695, USA

Abstract

Visual information from a speaker's face can enhance[1] or interfere with[2] accurate auditory perception. This integration of information across auditory and visual streams has been observed in functional imaging studies[3,4], and has typically been attributed to the frequency and robustness with which perceivers jointly encounter event-specific information from these two modalities[5]. Adding the tactile modality has long been considered a crucial next step in understanding multisensory integration. However, previous studies have 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[8–10]. Here we show that perceivers integrate naturalistic tactile information during auditory speech perception without previous training. Drawing on the observation that some speech sounds produce tiny bursts of aspiration (such as English 'p')[11], we applied slight, inaudible air puffs on participants' skin at one of two locations: the right hand or the neck. Syllables heard simultaneously with cutaneous air puffs were more likely to be heard as aspirated (for example, causing participants to mishear 'b' as 'p'). These results demonstrate that perceivers integrate event-relevant tactile information in auditory perception in much the same way as they do visual information.

Many languages use an expulsion of air, or 'aspiration', to convey basic lexical contrasts[12]. English speakers use this mechanism to distinguish aspirated sounds such as 'pa' and 'ta' from unaspirated sounds such as 'ba' and 'da'. All four human dermal mechanoreceptors[13], as well as hair-follicle mechanoreceptors[14], respond to air puffs. Aerodynamically, a puff is characterized as a short burst of turbulent airflow with a relatively higher initial pressure[15,16], typical of the transient pressure pattern produced in aspirated speech sounds[17].

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, matching for duration (390–450 ms each), fundamental frequency (falling pitch from 90 Hz to 70 Hz) and intensity (normalized to 70 decibels (10^{-5} W m^{-2})). Participants heard syllables in two separate blocks: one containing only labial consonants ('pa' and 'ba'), the other containing only alveolar consonants ('ta' and 'da'). The 16 unique tokens in each block were heard four 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 Supplementary Tables 1–3.

Correspondence and requests for materials should be addressed to B.G. (gick@interchange.ubc.ca).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions B.G. conceived and designed the experiment; D.D. designed and performed the data analysis.

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.

In our first experiment, air puffs were applied cutaneously on the dorsal surface of the hand between the right thumb and forefinger through ¼-inch (0.635-cm) vinyl tubing at 6 pounds per square inch (p.s.i.; 6 p.s.i. ≈ 421.84 g cm^{-2}) fixed at 8 cm from the skin surface. The back of the hand was chosen because it has high tactile sensitivity[18], and because it is a location where tactile stimulation including airflow has been observed to elicit non-specific activation of some second-stage auditory cortical neurons in macaques[19].

We considered that participants may have a good deal of previous experience with air puffs on the hand coupled with speech sounds—from concurrently hearing their own voice and feeling their own breath on their hands during speech. To determine whether the interaction would persist even at a body location lacking frequent self-experience, we designed a second experiment in which we applied air puffs to the centre of the neck at the suprasternal notch—a location where participants typically receive no direct airflow during their own speech production (though perceivers presumably do, at least on rare occasion, feel interlocutors’ aspirated air on their skin). As with the hand experiment, air puffs were delivered through ¼-inch vinyl tubing at 6 p.s.i. fixed at 8 cm from the skin surface.

In addition to the hand and neck trials, an ‘auditory-only’ experiment was designed to ensure that delivery of the air puffs was inaudible to participants. In this trial, the ¼-inch tube was placed immediately beside the participants’ right headphone at a distance of 5 cm and a pressure of 6 p.s.i., aimed tangentially forward so that airflow was not felt directly on the skin or hair.

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 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 before vowel onset and ending at the moment of vowel onset, thus simulating the timing of naturally produced English aspirated consonants.

Male and female participants were tested in all experiments. Before the experiment, participants were told that they might experience background noise and unexpected puffs of air. Participants were seated in a soundproof booth and asked to identify by pressing a button whether they heard ‘pa’ or ‘ba’ in the labial block, and ‘ta’ or ‘da’ in the alveolar block. Participants were then blindfolded and provided with auditory stimuli through sound-isolating headphones. The setup of equipment to deliver tactile stimuli was completed after the participants were blindfolded to conceal the body location of air puffs.

A mixed design repeated-measures analysis of variance was conducted with two consonant aspiration conditions (aspirated and unaspirated) by two airflow conditions (presence and absence) by two places of articulation (labial and alveolar) by three experiments (hand, neck and auditory-only). Results indicated weak main effects of aspiration ($F(1,63) = 5.426$, $P = 0.023$) (that is, perceivers identified unaspirated stops slightly more readily across all experiments) and place ($F(1,63) = 6.714$, $P = 0.012$) (that is, perceivers were slightly more accurate discerning alveolar versus labial stops), and strong main effects of aspiration × airflow ($F(1,63) = 26.095$, $P < 0.001$) (airflow caused perception of both unaspirated and aspirated stops as aspirated more often) and aspiration × airflow × experiment ($F(2,63) =$

7.600, $P = 0.001$) (that is, the effect of airflow applied to the neck and hand experiments, but not to the auditory-only experiment). There was no significant main effect of airflow, or of interaction between airflow and experiment (that is, application of airflow does not affect overall accuracy of perception of stimuli). No other significant effects were observed.

To identify whether there were significant interactions between aspiration and airflow in the hand and neck experiments, but not the auditory-only experiment, separate analyses of variance with repeated measures factors of aspiration (aspirated versus unaspirated) and air puffs (present versus absent) were conducted for both the alveolar and labial blocks of all experiments. Furthermore, to determine whether these interactions demonstrated augmentation of aspirated stop perception as well as interference with unaspirated stop perception, one-way repeated-measures analyses of variance comparing air puffs (present versus absent) were run separately for aspirated and unaspirated tokens.

Results for the hand experiment showed that the interaction of air puffs with the perception of aspiration was significant ($\alpha = 0.05$) for both the alveolar ($F(1,21) = 17.888$, $P < 0.001$, partial $\eta^2 = 46.0\%$) and labial ($F(1,21) = 14.785$, $P < 0.001$, partial $\eta^2 = 41.3\%$) blocks (Fig. 1). Further, the presence of an air puff enhanced correct identification of aspirated tokens (‘pa’ ($F(1,21) = 14.309$, $P = 0.001$, partial $\eta^2 = 40.5\%$) and ‘ta’ ($F(1,21) = 8.650$, $P = 0.008$, partial $\eta^2 = 29.2\%$)), and interfered with correct identification of unaspirated tokens (‘ba’ ($F(1,21) = 5.597$, $P = 0.028$, partial $\eta^2 = 21.0\%$) and ‘da’ ($F(1,21) = 16.979$, $P < 0.001$, partial $\eta^2 = 44.7\%$)).

Results for the neck experiment showed that the interaction of air puffs with the perception of aspiration was significant for both the alveolar ($F(1,21) = 5.486$, $P = 0.029$, partial $\eta^2 = 20.7\%$) and labial ($F(1,21) = 8.404$, $P = 0.009$, partial $\eta^2 = 28.6\%$) blocks (Fig. 2). Further, the presence of an air puff enhanced correct identification of aspirated tokens (‘pa’ ($F(1,21) = 7.140$, $P = 0.014$, partial $\eta^2 = 25.4\%$) and ‘ta’ ($F(1,21) = 6.020$, $P = 0.023$, partial $\eta^2 = 22.3\%$)) and showed a weak effect of interference with correct identification of unaspirated tokens (‘ba’ ($F(1,21) = 3.421$, $P = 0.078$, partial $\eta^2 = 14.0\%$) and ‘da’ ($F(1,21) = 1.291$, $P = 0.269$, partial $\eta^2 = 5.8\%$)).

No significant interaction between aspiration and air puffs was found for the auditory-only experiment (alveolar or labial block, $F(1,21) < 1$), confirming that participants could not hear the airflow or compressor activation (Fig. 3).

Our findings support the hypothesis that the human perceptual system integrates specific, event-relevant information across auditory and tactile modalities in much the same way as has been previously observed in auditory-visual coupling. This effect occurs in perceivers without previous training or awareness of the task, and at body locations where the effect is unlikely to be reinforced by frequent experience. These results complement recent work showing the involvement of the somatosensory system in speech perception[20], suggesting that the neural processing of speech is more broadly multimodal than previously believed. The methods used in this paper represent a model that will enable future functional imaging studies of passive audio-tactile and visuo-tactile integration, as well as behavioural studies of multi-sensory perception in previously untested populations, including infants and the blind. As these findings describe perceptual enhancement during passive perception, they imply possible future directions in audio and telecommunication applications and aids for the hearing impaired.

METHODS SUMMARY

Synthetic air puffs

The airflow device consisted of a 3-gallon (11.35-l) 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) connected to a Campbell Hausfeld MP513810 air filter, which reduced the sound volume conducted through the ¼-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 with 60 ms voice onset time for our speaker's average (mean) 'pa', and close to the range of voice onset time of 54–80 ms for English word-onset voiceless (aspirated) stops[12]. The output pressure of the synthesized puffs was adjusted so that impact was minimally perceptible by participants. As such, microphone recordings at 8 cm showed an average peak relative non-dimensional pressure of 0.023 for the synthetic puffs, compared with 0.096 for our speaker's average 'pa'.

Procedure

In total, we tested 66 participants, 22 for each of the experimental trials (hand and neck) and the auditory-only trial. Half received the labial ('pa', 'ba') block first, and half received the alveolar ('ta', 'da') block first. Within each block, participants heard 12 practice tokens (six with and six without air puffs) followed by 16 experimental tokens for each condition (aspirated versus unaspirated, puff versus no puff, randomized), totalling 64 experimental tokens per block. A custom-built computer program written in Java 1.6 recorded responses from a customized keypad and presented new tokens 1,500 ms after each response. Half of the participants pressed the left button to indicate an aspirated response, and half pressed the right button.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Council of Canada to B.G., and by a National Institutes of Health Grant DC-02717 to Haskins Laboratories. We acknowledge the contributions of Y. Ikegami for initial piloting, R. Lee for piloting tapping protocols, L. McLeod for help setting up the University of British Columbia system, G. Ramsay and J. Wang for building switch mechanisms, S. Renardy for organizing protocols, D. H. Whalen for discussion and students of the University of British Columbia's Interdisciplinary Speech Research Laboratory for organizing participants and other assistance.

References

1. Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 1954; 26:212–215.
2. McGurk H, MacDonald JW. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]
3. Calvert GA, et al. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]
4. Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. 2003; 15:57–70. [PubMed: 12590843]
5. Diehl RL, Kluender KR. On the objects of speech perception. Ecol. Psychol. 1989; 1:121–144.
6. Fowler C, Dekle D. Listening with eye and hand: crossmodal contributions to speech perception. J. Exp. Psychol. Hum. Percept. Perform. 1991; 17:816–828. [PubMed: 1834793]

7. Gick B, Jóhannsdóttir KM, Gibraiel D, Mühlbauer J. Tactile enhancement of auditory and visual speech perception in untrained perceivers. J. Acoust. Soc. Am. 2008; 123:EL72–EL76. [PubMed: 18396924]
8. 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; 63:246–257. [PubMed: 632417]
9. Reed CM, Durlach NI, Braida LD, Schultz MC. Analytic study of the Tadoma method: effects of hand position on segmental speech perception. J. Speech Hear. Res. 1989; 32:921–929. [PubMed: 2601321]
10. 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]
11. Derrick D, Anderson P, Gick B, Green S. Characteristics of air puffs produced in English 'pa': data and simulation. J. Acoust. Soc. Am. 2009; 125:2272–2281. [PubMed: 19354402]
12. Lisker L, Abramson AS. A cross-language study of voicing in initial stops: acoustical measurements. Word. 1964; 20:384–423.
13. Mizobuchi K, et al. Single unit responses of human cutaneous mechanoreceptors to air-puff stimulation. Clin. Neurophysiol. 2000; 111:1577–1581. [PubMed: 10964067]
14. Sabah NH. Controlled stimulation for hair follicle receptors. J. Appl. Physiol. 1974; 36:256–257. [PubMed: 4811387]
15. Sangras R, Kwon OC, Faeth GM. Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids. J. Heat Transfer. 2002; 124:460–469.
16. Diez FJ, Sangras R, Kwon OC, Faeth GM. Erratum: Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids. *ASME J. Heat Transfer*, **124**, 460–469 (2002). J. Heat Transfer. 2003; 125:204–205.
17. Stevens, KN. Acoustic Phonetics. Vol. Ch. 7. MIT Press; 1998.
18. Weinstein, S. The Skin Sense. Kenshalo, DR., editor. Thomas; 1968. p. 195-222.
19. Fu K-MG, et al. Auditory cortical neurons respond to somatosensory stimulation. J. Neurosci. 2003; 23:7510–7515. [PubMed: 12930789]
20. Ito T, Tiede M, Ostry DJ. Somatosensory function in speech perception. Proc. Natl Acad. Sci. USA. 2009; 106:1245–1248. [PubMed: 19164569]

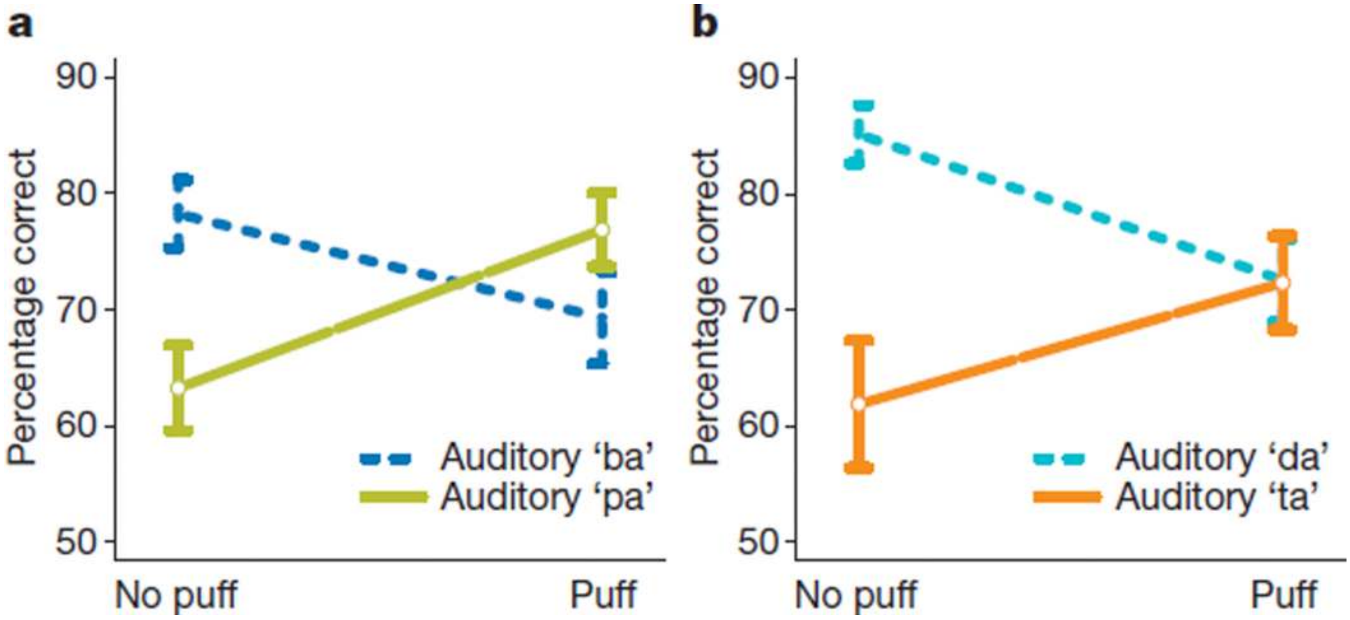

Figure 1. Interaction graphs for the hand experiment with standard error bars

a, Labial; **b**, alveolar.

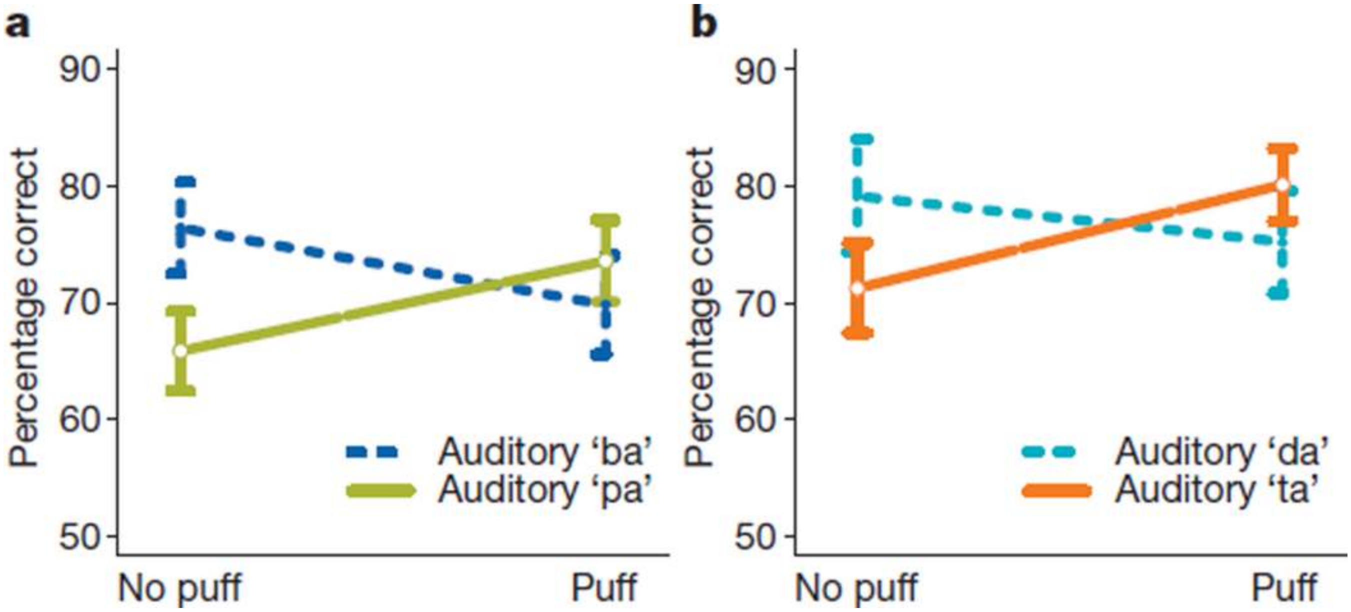

Figure 2. Interaction graphs for the neck experiment with standard error bars
a, Labial; **b**, alveolar.

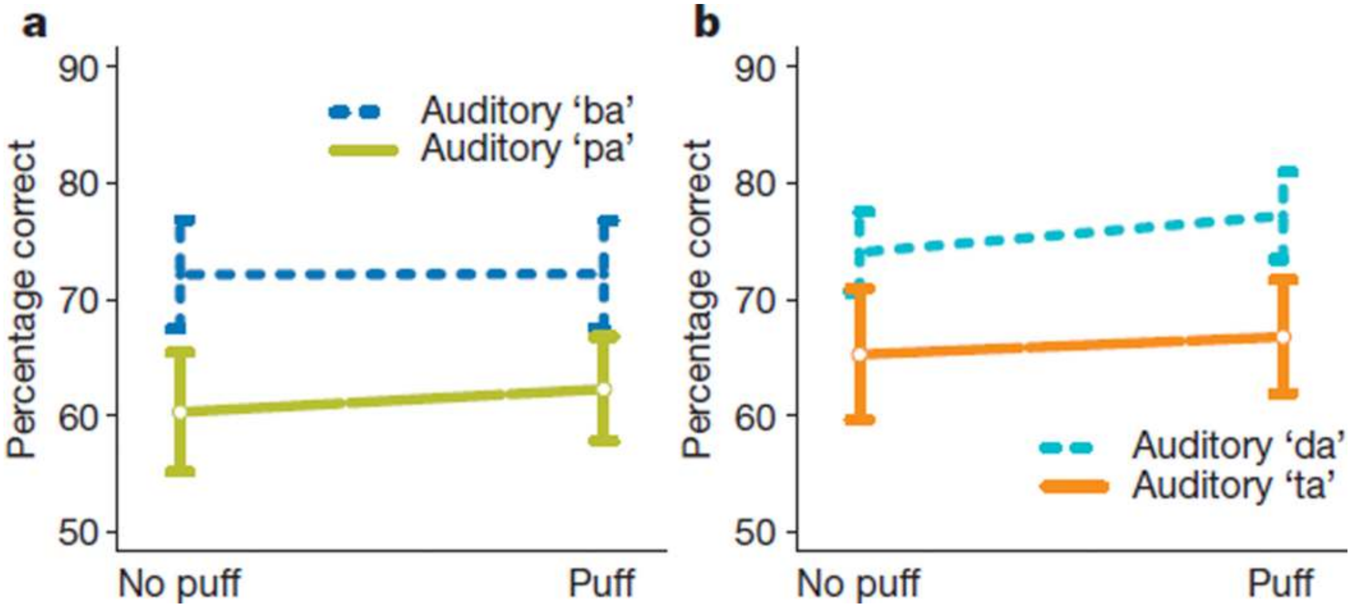

Figure 3. Interaction graph for control experiment with standard error bars
a, Labial; **b**, alveolar.