AERO-TACTILE INFLUENCE ON SPEECH PERCEPTION OF VOICING CONTINUA

Dolly Goldenberg[1,2], Mark K. Tiede[2], D. H. Whalen[1,2,3]

[1]Yale University [2]Haskins Laboratories [3]City University of New York
dolly.goldenberg@yale.edu tiede@haskins.yale.edu dwhalen@gc.cuny.edu

ABSTRACT

Previous work [9, 7] has established that puffs of air applied to the skin and timed with listening tasks bias the perception of voicing by naive listeners. The current study has replicated and extended these findings by testing air puff effects on gradations of a voicing continuum rather than the voiced and voiceless exemplars of the original work. This design has the advantage of distinguishing responses differentially along the continuum, and our results show that while overall coincident air puffs bias listener judgments towards increased voiceless responses, the effects are least at the continuum endpoints. This suggests that during integration auditory and aerotactile inputs are weighted differently by the perceptual system, with the latter exerting greater influence in cases where the auditory cues to voicing are ambiguous.

Keywords: multimodal speech perception, multimodal integration

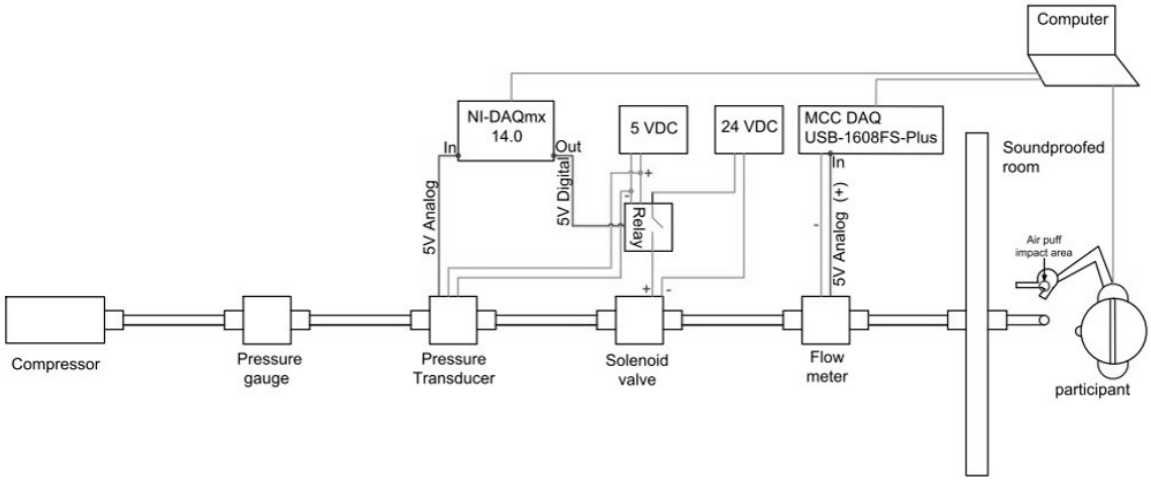
1. INTRODUCTION

Integration of different sensory modalities in speech perception is well documented, especially in the audio-visual domain. It has been shown that visual cues enhance auditory speech perception in both suboptimal acoustic conditions such as background noise or heavy foreign accent [23, 19, 14] and in cases of increased cognitive load such as complicated structure or content [19, 1]. Conversely, incongruent visual cues have also been shown to interfere with auditory perception in adults [17, 16] and infants [4, 20].

Recently, evidence for audio-tactile integration has also been presented. [12] showed that by using a robotic device to create patterns of facial skin deformation in listeners that would normally accompany production, perceptual judgments of a vowel continuum were shifted correspondingly. Tactile effects on perception have been demonstrated for participants with explicit knowledge of the task [8, 10], or when trained to make a connection between the tactile and the auditory cues [18, 3, 21]. Recent studies have also established the effects of tactile information on auditory perception for uninformed and untrained listeners [9, 7]. Specifically, [9] studied the effect of applying air puffs to the back of the hand or the center of the neck on the perception of voiced and voiceless CV exemplars such as /pa/ and /ba/ in background noise. They found that the presence of airflow affected both the identification of aspirated stops by speakers of English, by enhancing it, and the identification of unaspirated stops by speakers of English, by interfering with it. Since no such effect was found for the participants in the control group, where no direct tactile information was provided, they concluded that speech is truly a multisensory phenomenon, and that tactile information can modulate speech perception similar to the way vision does.

The present study aims to further investigate the nature of audio - aerotactile integration. Evidence for multimodal speech perception, both audio-visual and audio-tactile, has served as an argument for the independence of the objects of speech perception from the units of non-speech auditory perception (See [24, 11] for discussion). The argument is that since the objects of speech perception are multimodally accessible while sounds are not, sounds cannot be the objects of speech perception. This argument relies crucially on the interpretation of experimental findings as supporting multisensory integration. However, in the work presented by [9, 7], it has been claimed that this interpretation is not sufficiently supported by the data. For instance, [15] argues that it is possible that the participants used the airflow, when it was provided, as their sole cue, and thus interpreted the stimuli based only on tactile information without integration with the auditory modality. Accordingly, one goal of the current study is to address this critique. Instead of the unambiguously voiced or unvoiced stimuli used in [9] we make use of a voicing continuum, including sounds that span the range between aspirated and unaspirated. Evidence for multisensory integration can then be assessed on whether judgments of such

Figure 1: Overview of the aero-tactile stimulus presentation system



intermediate sounds are systematically affected by aero-tactile cues.

To examine whether air puff effects are related to the salience of aspiration as a production cue we extend testing for the first time to two additional continua, the first a voicing contrast at the velar place of articulation, and the second an unrelated (vowel quality) continuum. We expect that given the smaller aspiration cues associated with velars [22] the effects of air puffs will be correspondingly less, and that as aspiration is irrelevant in the vowel quality distinction no effects of air puffs will be observed in that case.

2. METHOD

2.1. Participants

Eighteen monolingual right-handed native speakers of American English (ten females, eight males) participated in the experiment. The participants were naive to the purpose of the study and had no self-reported speech or hearing deficiencies. They were compensated for their time.

2.2. Auditory Stimuli

The stimuli were created by recording a male monolingual native speaker of American English producing six repetitions of the syllables /pa/, /ba/, /ka/, and /ga/. A second speaker produced target /hɪd/ and /hɛd/ for construction of the vowel quality continuum. Two eight-step voicing continua were created, one for each place of articulation. The continua were created by removing the release burst from a voiceless token and then shortening the aspiration in eight log-scaled steps. A viability test was run with an independent group of participants that did not take part in the study (n = 41) in order to assess the quality of the stimuli. The participants were asked to choose whether they heard a voiceless or voiced token, and to rate the goodness of the token. The sounds from the two continua were presented in random order. The perceived category boundary was established by the *bias*, that is, the 50% crossover point of the psychometric function for each continuum computed across all listeners, and the sharpness of the boundary by its *acuity* (slope). The labial category boundary was found to be approximately centered (bias = 4.2, acuity = 1.1), while the velar category boundary was skewed in the direction of voiceless responses (bias = 3.6, acuity = 2.0). The goodness ratings were higher at the extremes than at the medial steps of each continuum, which reflects the fact that the ambiguous sounds were harder to categorize, as expected.

2.3. Aero-Tactile Stimuli

Air puff stimuli were produced by an air compressor using a programmable solenoid valve providing constant flow, delivered by tubing inserted into a soundproofed room through a cable port (see Figure 1 for a schematic description of the system). A custom Matlab (MathWorks) procedure was developed to synchronize the opening of the air valve with playback of the auditory stimulus. The durations of the delivered puffs were 87 ms for the labial condition and 92 ms for the velar condition, reflecting the mean voice onset times (VOTs) of the six voiceless tokens of each type produced by the speaker, and which fall within the VOT range of initial aspirated stops in American English (54-100 ms, [13, 5]). Puff durations for the vowel quality con-

trol condition were 92 ms. Flow rate while active was constrained by a flowmeter to be a uniform 5 Standard Liters Per Minute (SLPM). The exit point of the airtube was placed 5 cm away from the dorsal surface of the hand between the right thumb and forefinger, creating an area of initial impact with a diameter of 2-3 cm [6].

2.4. Procedure

The experiment included two parts, an initial test to verify that the air puffs were felt but not heard, seen or otherwise perceived, and the main part, which tested participant responses to the auditory stimuli in the presence and absence of air puffs. Audio stimuli were presented through headphones, and a small fan provided background noise to mask potential cues associated with puff release through the tube.

2.4.1. Puff Detection Test

For each trial in the initial test participants heard a one second 500 Hz tone presented through headphones which was followed by a 50 ms long air puff in 50% of the trials. Air puff presentation (present vs. absent) was randomized. The test had two blocks of 50 trials each. During the first block the air tube was positioned such that the participants felt the puffs on their hand, and during the second block, it was positioned such that they could not feel the air flow. In each trial participants were asked to indicate using a two button keypad labeled 'Y' (yes) and 'N' (no) whether they felt or otherwise detected a puff or not.

2.4.2. Perturbed Continua Testing

For the main experiment, each participant's hand was positioned as described above, such that they could consistently feel the puff of air on the back of their hand when present. Trials were organized into blocks, in which stimuli were drawn consistently from one of the three continua tested: labial (/pa/ to /ba/), velar (/ka/ to /ga/), and vowel quality (/hɛd/ to /hɪd/). Each block included six repetitions of each of the eight stimuli constructed for that continuum. Air puffs were presented during three of the six repetitions. Within an experiment each participant received five blocks each of two different continuum types, resulting in three repetitions x two puff types (+/-) x eight continuum steps x five blocks for a total of 240 separate judgments per continuum type, with 15 per condition at each continuum step. Twelve of the participants heard five velar blocks and five labial blocks, Three of the participants heard five velar blocks and five vowel blocks, one participant heard five labial blocks and five vowel blocks, and two participants heard five labial blocks only. In each trial participants were asked to indicate the sound they heard using a two button keypad labeled appropriately for the continuum ('P' or 'B', 'K' or 'G', 'hid' or 'head'). The stimulus presentation order was randomized within blocks, and blocks alternated between continuum type. New tokens were presented 1,000 ms after each response. The same procedure that controlled stimulus presentation also recorded participant responses from the response keypad, together with airflow as transduced by the flow meter, and at the exit point of the tube, by a directional microphone. The auditory stimuli the participants heard were recorded as well to verify puff presentation timings. In a few instances an air puff was scheduled but not properly delivered, or unscheduled but inappropriately delivered, and these trials were excluded.

3. RESULTS

3.1. Puff Detection Test

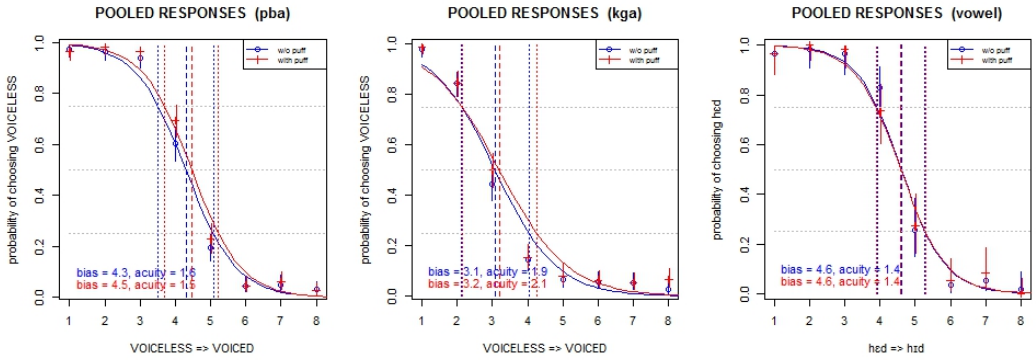
The participants were able to recognize the puffs 95% of the time during the first block (worst performer 76%), when their hand was close to the exit point of the tube. When their hand was completely removed from the exit point of the tube, they only recognized the puffs 50% of the time (best performer 54%). An exact binomial test confirms that in the first case, the recognition percentages were well above chance, while in the second they were not. These results confirm that participants were able to consistently feel the puff of air on their hand, but could not hear, see, or otherwise detect it.

3.2. Perturbed Continua Testing

Figure 2 shows the perceived category boundary, pooled across speakers, in the presence and absence of air puffs. The y axis represents the probability of choosing a voiceless token or /hɛd/. The x axis represents values at each of the eight steps along the continuum. The fitted lines show the estimated psychometric function. The baseline condition, presented without puffs, is shown in black, and the perturbation condition with air puffs is shown in gray. The shift of the bias to the right in the presence of air puffs in the voicing continua reflects the fact that there were more voiceless responses in this condition.

A generalized linear mixed-effects model[1] computed with the lme4 package in R [2] was used to

Figure 2: Perceived category boundaries pooled across speakers with (gray) and without (black) air puff perturbation. Responses at each continuum step show 95% C.I.s and define the psychometric function estimates. The left panel shows the labial (/pa/:/ba/) continuum, the center panel the velar (/ka/:/ga/) continuum, and the right panel the vowel (/hɛd/ to /hɪd/) continuum.



predict the response for each continuum given the fixed effect of presence/absence of air puffs, with random intercepts for continuum step nested within the participant ID. In the labial block, the presence of air puffs was found to increase the likelihood of voiceless response. In the velar block, the presence of air puff was found to marginally increase the likelihood of voiceless response. In the vowel blocks, the presence of air puffs had no effect on the responses (see Table 1 for details of the model output).

Table 1: GLMM response model

	-Air puff (baseline) vs. +Air puff				
continuum	**coef**	**z**	**p**	**sig.**	**dir.**
Labial	0.293	2.448	0.014	*	+ > -
Velar	0.261	1.927	0.054	.	+ > -
Vowel	-0.014	-0.058	0.953	n.s	

Further investigation of the bias point for the response function of the labial block revealed that the effect was not homogeneous across the continuum: the bias point for the pooled responses with air puffs (4.5) is higher than the bias point for the pooled responses without air puffs (4.3). That is, for steps 1:5 (roughly the voiceless end of the continuum) the presence of air puffs enhanced the likelihood of perceiving a voiceless sound. For steps 6:8, however, the air puffs did not appear to affect the perceivers' judgments. While this finding may reflect the fact that our voiced stimuli were constructed from a voiceless token and may lack some of the cues for voicing, it may also suggest that the air puffs were more effective as enhancement of perception of voiceless sounds, and less effective as interference with perception of voiced sounds.

4. DISCUSSION

These results replicate the findings presented in [9] by showing that aero-tactile information does affect perception of voicing. We extend their results in three ways. First, our results use a within- rather than between-groups design, such that each participant served as their own control. Second, our stimuli included intermediate and potentially ambiguous steps along a continuum rather than endpoints masked by white noise alone, and our results show that aero-tactile perturbation had its greatest effect on these intermediate tokens. Third, we show that the perturbatory effect is smaller for the velar continuum, and nonexistent for the vowel quality continuum, as expected given the smaller (for velar) and irrelevant (for vowels) role of aspiration in producing those contrasts.

The reduction of the effectiveness of the puff for the velars may reflect detailed knowledge of articulation, namely, that the air must travel farther through the mouth for velars than for labials. The fact that the voiceless end of the continuum was affected by the air puffs more than the voiced end of it suggests that this aero-tactile information does not successfully compete with the auditory cues associated with voiced sounds. However, it is possible that reaction times might show effects even in the face of overriding information [25]. Overall, these findings provide support for multimodal integration in speech perception, as they show consistent effects tied to the salience of aspiration in production of the relevant sounds. This weakens the argument of [16] since it suggests that cues are interpreted not uniformly but rather with respect to relevance, and that aero-tactile information (along with auditory information) is involved in this process.

Acknowledgements

This research was funded by National Institutes of Health (NIH) Grant DC-002717 to Haskins Laboratories.

5. REFERENCES

[1] Arnold, P., Hill, F. 2001. Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology* 92(2), 339–355.

[2] Bates, D., Maechler, M., Bolker, B. M., Walker, S. 2014. lme4: Linear mixed–effects models using Eigen and S4. ArXiv e-print; submitted to Journal of Statistical Software, http://arxiv.org/abs/1406.5823.

[3] Bernstein, L. E., Demorest, M. E., Coulter, D. C., Oćonnell, M. P. 1991. Lipreading sentences with vibrotactile vocoders: Performance of normal–hearing and hearing–impaired subjects. *The Journal of the Acoustical Society of America* 90(6), 2971–2984.

[4] Burnham, D., Dodd, B. E. 1996. Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In: Stork, D. G., Hennecke, M. E., (eds), *Speech reading by Humans and Machines: Models, Systems, and Applications*. Berlin: Springer–Verlag 103–114.

[5] Cooper, A. M. 1991. *An articulatory account of aspiration in English*. PhD thesis Yale University.

[6] Derrick, D., Anderson, P., Gick, B., Green, S. 2009. Characteristics of air puffs produced in English "pa": Experiments and simulations. *The Journal of the Acoustical Society of America* 125(4), 2272–2281.

[7] Derrick, D., Gick, B. 2013. Aerotactile integration from distal skin stimuli. *Multisensory Research* 26, 405–416.

[8] Fowler, C. A., Dekle, D. J. 1991. Listening with eye and hand: Cross–modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 17(3), 816–828.

[9] Gick, B., Derrick, D. 2009. Aero-tactile integration in speech perception. *Nature* 462, 502–504.

[10] Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., Mühlbauer, J. 2008. Tactile enhancement of auditory and visual speech perception in untrained perceivers. *The Journal of the Acoustical Society of America* 123(4), EL72–EL76.

[11] Goldstein, L. M., Fowler, C. A. 2003. Articulatory phonology: A phonology for public language use. In: Schiller, N., Meyer, A., (eds), *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter 159–207.

[12] Ito, T., Tiede, M., Ostry, D. J. 2009. Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the united States of America* 106(4), 1245–1248.

[13] Lisker, L., Abramson, A. S. 1967. Some effects of context on voice onset time in english stops. *Language and Speech* 10(1), 1–28.

[14] Macleod, A., Summerfield, Q. 1990. A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology* 24(1), 29–43.

[15] Massaro, D. W. 2009. Caveat emptor: The meaning of perception and integration in speech perception. Available from Nature Precedings http://hdl.handle.net/10101/npre.2009.4016.1.

[16] Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., Tsuzaki, M. 1993. Bimodal speech perception: An examination across languages. *Journal of Phonetics* 21(4), 445–478.

[17] McGurk, H., MacDonald, J. 1976. Hearing lips and seeing voices. *Nature* 264, 746–748.

[18] Reed, C. M., Durlach, N. I., Braida, L. D., Schultz, M. C. 1989. Analytic study of the Tadoma Method: Effects of hand position on segmental speech perception. *Journal of Speech, Language, and Hearing Research* 32, 921–929.

[19] Reisberg, D., McLean, J., Goldfield, A. 1987. Easy to hear but hard to understand: A lip–reading advantage with intact auditory stimuli. In: Dodd, B. E., Campbell, R., (eds), *Hearing by eye: The psychology of lip–reading*. Hillsdale, NJ: Lawrence Erlbaum 97–114.

[20] Rosenblum, L. D., Schmuckler, M. A., Johnson, J. A. 1997. The McGurk effect in infants. *Attention, Perception, and Psychophysics* 59(3), 347–357.

[21] Sparks, D. W., Kuhl, P. K., Edmonds, A. E., Gray, G. P. 1978. Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of segmental features of speech. *The Journal of the Acoustical Society of America* 63(1), 246–257.

[22] Stevens, K. N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

[23] Sumby, W. H., Pollack, I. 1954. Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America* 26(2), 212–215.

[24] Trout, J. D. 2001. The biological basis of speech: What to infer from talking to the animals. *Psychological Review* 108(3), 523–549.

[25] Whalen, D. H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics* 35(1), 49–64.

[1] glmer (RESP∼PUFF + (1|ID/STEP), family=binomial)