

# Speech sounds alter facial skin sensation

Takayuki Ito and David J. Ostry

*J Neurophysiol* 107:442-447, 2012. First published 19 October 2011; doi:10.1152/jn.00029.2011

---

## You might find this additional info useful...

---

This article cites 30 articles, 7 of which can be accessed free at:

<http://jn.physiology.org/content/107/1/442.full.html#ref-list-1>

Updated information and services including high resolution figures, can be found at:

<http://jn.physiology.org/content/107/1/442.full.html>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

---

This information is current as of January 4, 2012.

## Speech sounds alter facial skin sensation

Takayuki Ito<sup>1</sup> and David J. Ostry<sup>1,2</sup>

<sup>1</sup>Haskins Laboratories, New Haven, Connecticut; and <sup>2</sup>Department of Psychology, McGill University, Montréal, Québec, Canada

Submitted 12 January 2011; accepted in final form 19 October 2011

**Ito T, Ostry DJ.** Speech sounds alter facial skin sensation. *J Neurophysiol* 107: 442–447, 2012. First published October 19, 2011; doi:10.1152/jn.00029.2011.—Interactions between auditory and somatosensory information are relevant to the neural processing of speech since speech processes and certainly speech production involves both auditory information and inputs that arise from the muscles and tissues of the vocal tract. We previously demonstrated that somatosensory inputs associated with facial skin deformation alter the perceptual processing of speech sounds. We show here that the reverse is also true, that speech sounds alter the perception of facial somatosensory inputs. As a somatosensory task, we used a robotic device to create patterns of facial skin deformation that would normally accompany speech production. We found that the perception of the facial skin deformation was altered by speech sounds in a manner that reflects the way in which auditory and somatosensory effects are linked in speech production. The modulation of orofacial somatosensory processing by auditory inputs was specific to speech and likewise to facial skin deformation. Somatosensory judgments were not affected when the skin deformation was delivered to the forearm or palm or when the facial skin deformation accompanied nonspeech sounds. The perceptual modulation that we observed in conjunction with speech sounds shows that speech sounds specifically affect neural processing in the facial somatosensory system and suggest the involvement of the somatosensory system in both the production and perceptual processing of speech.

multisensory integration; auditory-somatosensory interaction; cutaneous perception; speech perception

EVIDENCE FOR MULTISENSORY INTEGRATION in speech perception is generally centered on its effects on auditory function. Visual information facilitates auditory perception (Sumbly and Pollack 1956) and results in illusory auditory percepts such as the McGurk effect (McGurk and MacDonald 1974). Somatosensory inputs likewise affect the auditory perception of speech sounds (Gick and Derrick 2009; Ito et al. 2009). But does the multisensory processing of speech sounds specifically target the auditory system, as might be expected if speech was primarily an auditory specialization (Diehl et al. 2004)? Or are the above examples just instances of a more broadly based multisensory interaction (Liberman et al. 1967)?

Interactions between the auditory and somatosensory systems have been documented previously in nonspeech contexts such as the detection of events in different spatial locations (Murray et al. 2005; Tajadura-Jiménez et al. 2009) and temporal frequency discrimination (Yau et al. 2009, 2010). The associated neural mechanisms have been investigated using brain imaging in humans (Beauchamp et al. 2008; Foxe et al. 2002; Murray et al. 2005) and in other primates (Fu et al. 2003;

Kayser et al. 2005; Lakatos et al. 2007). The effects of auditory inputs on somatosensory perception have been reported in cases other than speech. For example, the detection of near-threshold somatosensory stimuli is improved by the presence of simultaneous auditory inputs (Ro et al. 2009). Similarly, the perception of surface roughness is affected by sounds played to subjects while they rub their hands together (Jousmaki and Hari 1998). Judgments of the crispness of potato chips are dependent on the intensity and frequency of sounds played as subjects bite on the chips (Zampini and Spence 2004). Auditory and somatosensory systems also interact in speech perception. In particular, facial somatosensory inputs similar to those that arise in the production of speech have been found to systematically alter the perception of speech sounds (Ito et al. 2009). If speech perception mechanisms are tightly linked to those of speech production, speech-specific auditory-somatosensory interactions should be observed in the reverse direction, that is, speech sounds should modulate the perception of somatosensory inputs that would normally be associated with speech articulatory motion.

We tested the idea that speech sounds affect facial somatosensory perception by using a robotic device to generate patterns of facial skin deformation that are similar in timing and duration to those experienced in speech production. We combined the skin stretch with the simultaneous presentation of speech sounds. We found that skin stretch judgments are affected by auditory input and that the effects are, for the most part specific, to speech sounds and facial skin sensation. Auditory inputs must be speechlike to affect somatosensory processing, and the auditory influence on the somatosensory system is limited to somatosensory inputs that would normally accompany the production of speech. Taken together with our complementary finding (Ito et al. 2009), the results underscore the reciprocal interaction between the production and perception of speech.

### METHODS

**Participants and sensory test.** Fifty-six native speakers of American English participated in the experiment. Participants were all healthy young adults with normal hearing. All participants signed informed consent forms approved by the Human Investigation Committee of Yale University.

We asked participants to indicate which of two sequential stretches (see Fig. 1A as an example) was greater in amplitude by pressing buttons on a computer keyboard. The test was carried out with participants' eyes covered. The auditory stimuli were the computer-generated sounds of the words "head" and "had" (Ito et al. 2009). The stimuli were presented one at a time through headphones and were timed to coincide with the skin stretch to the cheeks. The relative timing of audio and somatosensory stimulation was the same as in a previous study (Ito et al. 2009), so that participants perceived the

Address for reprint requests and other correspondence: D. J. Ostry, Dept. of Psychology, McGill Univ., 1205 Dr. Penfield Ave., Montréal, QC, Canada H3A 1B1 (e-mail: david.ostry@mcgill.ca).

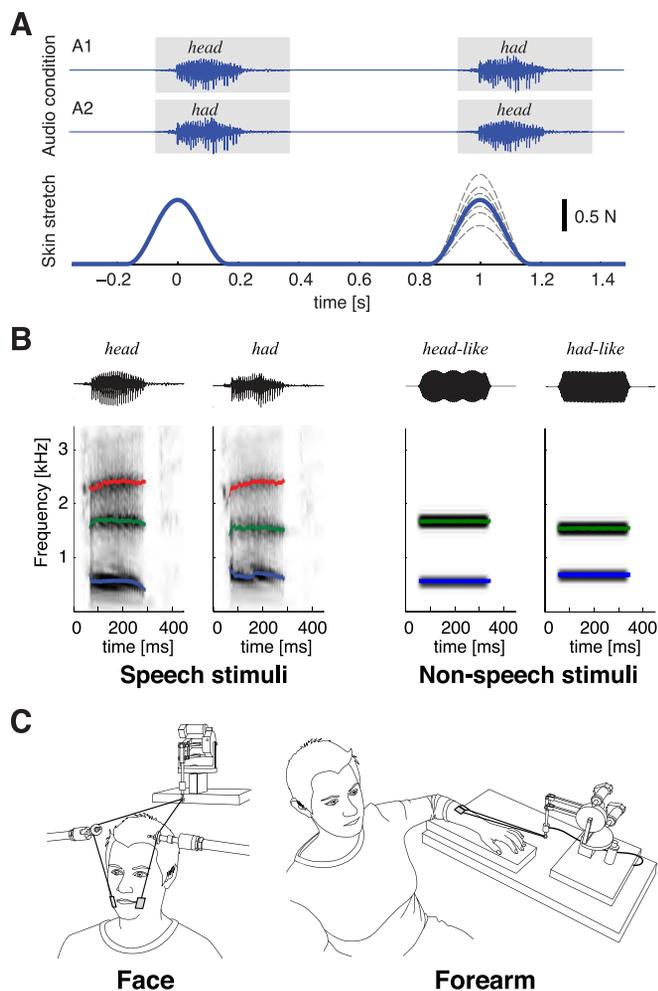


Fig. 1. Experimental setup. *A*: patterns of auditory and somatosensory stimulation in *experiment 2*. *Top*, patterns of auditory stimulation [labeled as A1 (“head-had”) and A2 (“had-head”)]; *bottom*, different force profiles of skin stretch. The solid line shows the condition in which the two sequential stretches are equal in magnitude. The dotted line shows all other variations. *B*: spectrograms showing audio stimuli. *Left*, speech stimuli (“head” and “had”); *right*, nonspeech stimuli (mixed pure tones). *C*: illustration of stimulation sites and stimulus direction. *Left*, facial skin stretch; *right*, site on the forearm.

inputs as simultaneous. We experimentally manipulated the order of the speech sounds. In one condition (A1), the word “head” was presented with the first skin stretch, and the word “had” was presented with the second skin stretch. In the other condition (A2), the opposite order was used (see Fig. 1A).

**Experimental manipulation.** We programmed a small robotic device to apply skin stretch loads. The skin stretch was produced using two small plastic tabs ( $2 \times 3$  cm each) attached bilaterally with tape to the skin at the sides of the mouth. The skin stretch was applied upward (Fig. 1C).

We examined the effects of hearing the words “head” and “had” on judgments of skin stretch magnitude. We chose these particular utterances because of the movement sequence involved in their production. Both utterances involve a simple pattern of vocal tract opening and closing. Both begin from a neutral start position for the sound /h/. The utterances differ in terms of the maximum aperture for the sounds /ɛ/ in “head” and /æ/ in “had,” such that the vocal tract opening is greater for “had.” They share a common end position for the final midpalatal stop consonant, /d/. The sounds for “head” and “had” were computer synthesized by shifting the first (F1) and second (F2) formant frequencies from values observed for “head” to those

associated with “had.” Apart from the differences in F1 and F2, the acoustic characteristics (duration, mean intensity, and third and higher formant frequencies) of the two utterances were the same. The details have been described in a previous publication (Ito et al. 2009). We reasoned that even though the stimuli were auditory in nature, they would influence judgments of applied force in a manner that related to the characteristics of their production. Thus, since production of the word “had” involves larger-amplitude tongue and jaw movement than “head,” somatosensory input due to skin stretch would be more heavily weighted by the sound “had” than the sound “head.” Our overall assumption was that the perceptual processing of speech inputs includes a fundamental somatosensory component.

**Experiment 1** compared somatosensory perceptual judgments under control (CTL) conditions (skin stretch in the absence of speech) with those obtained when speech sounds and skin stretch occurred simultaneously. Eight individuals participated in three experimental conditions: two involving speech sounds, “head-had” (A1) and “had-head” (A2), and one involving a no-sound condition (CTL). Speech sound trials alternated with no-sound trials. We presented speech-shaped noise in the background throughout the entire period of the experiment. The noise level was 20 dB less than that of the speech stimuli. Skin stretch stimuli were delivered twice on each trial, with an interval of 1,300 ms between the two stretches. We used two amplitudes of skin stretch, either 1.0 or 1.2 N, and three combinations of stimuli were used: 1) first and second stimuli were both 1.0 N, 2) the first stimulus was 1.2 N and the second stimulus was 1.0 N, and 3) the first stimulus was 1.0 N and the second stimulus was 1.2 N. Of these three combinations, *condition 1* is the condition of greatest interest since it permitted us to directly assess whether somatosensory perception is modified by auditory input. *Conditions 2* and *3* were included to ensure that subjects felt that they could detect differences in force levels at least some of the time. Force combinations were presented in random order. Thirty responses were recorded for each somatosensory-audio condition.

When judgments of skin stretch magnitude were obtained with unequal force levels, subjects performed at a uniformly high level. Subjects correctly judged a 1.2-N stretch as greater than a 1-N stretch in  $88 \pm 0.21\%$  (mean  $\pm$  SE) of the cases. As with equal stretches (see above), the word “head” reduced the probability that a stretch would be judged greater and the word “had” increased it. However, the effects were smaller than those observed for equal stretch magnitudes, and differences were not statistically reliable, presumably because judgment accuracy was near to perfect under these conditions.

**Experiment 2** examined the effects of speech sounds on somatosensory function over a wider range of force differences. It also involved three CTL tests that assessed the extent whether the effect of auditory input on somatosensory function is specific to speech sounds and speechlike somatosensory inputs. Twelve different individuals participated in each of the four tests (48 subjects in total). We used a between-subjects design to avoid the possibility that auditory stimuli in nonspeech conditions might be perceived as speech (or speechlike) because of prior experience with corresponding speech stimuli (Remez et al. 1981). In the somatosensory perceptual test, the first of two sequential forces was constant at 1.0 N. The amplitude of the second force was set at one of seven force levels relative to the first force ( $-0.4, -0.2, -0.1, 0, 0.1, 0.2, \text{ and } 0.4$  N; dashed lines in Fig. 1A). The peak-to-peak interval between the two commanded forces was 1,000 ms. We used the two same audio conditions as in *experiment 1* (A1: “head-had” and A2: “had-head”). In total, there were 14 somatosensory-audio conditions (7 somatosensory conditions  $\times$  2 auditory conditions), which were tested in random order. We recorded 26 responses for each somatosensory-audio combination.

CTL tests examined the effects of speech versus nonspeech auditory stimuli and also the effects of facial versus nonfacial skin stretch on somatosensory perception. In one CTL test, we used nonspeech sounds matched in frequency to the first two formants of the speech stimulus. We used these two formants specifically because they are the

primary acoustical measures that distinguish “head” and “had.” The nonspeech sounds were produced using pure tones (see Fig. 1B). The frequencies of two pure tones were set to match the F1 and F2 of /e/ in “head” (F1: 566 Hz and F2: 1,686 Hz) and the F1 and F2 of /æ/ in “had” (F1: 686 Hz and F2: 1,552 Hz), respectively. The duration and timing of stimulus presentation matched those of the vowels of the original speech stimuli. The somatosensory stimuli used for this CTL test were the same as in the primary manipulation. In the second CTL test, the skin stretch was applied to the hairy skin of the forearm instead of the facial skin (see Fig. 1C). The skin stretch was applied in the direction of the hand. The auditory stimuli were “head” and “had,” as in the primary manipulation. In the third CTL test, we applied the same somatosensory stimuli to the hairless skin of the palm, where the cutaneous receptor density is higher. Two of these subjects were subsequently excluded as their scores fell beyond 2 SDs from the mean (one in each direction). All experiments were carried out with the subjects’ eyes covered. As in *experiment 1*, subjects were required to indicate whether the first or second skin stretch felt greater in amplitude.

**Statistical analysis.** The probability that subjects judged the second skin stretch as larger than the first was the dependent measure. In *experiment 1*, we compared means for the somatosensory judgments in the no-sound condition with those of the “head-had” (A1) and “had-head” (A2) conditions. Performance was quantified on a per subject basis, with judgment probabilities in the three conditions converted to *z*-scores. We conducted one-way repeated-measures ANOVA in which the independent variable was the auditory test condition [three levels: A1 (“had-head”), A2 (“head-had”), and no audio]. Post hoc contrasts were carried out using Tukey tests.

In *experiment 2*, the probability that the participant judged the second force as greater was calculated for each of the seven somatosensory conditions. We obtained fitted psychometric functions from the associated judgment probabilities by carrying out logistic regression using a generalized linear model (Hosmer and Lemeshow 2000) with two independent variables (somatosensory force difference and auditory presentation order). The identification crossover was obtained from the regression by finding the difference between the two skin stretch forces that corresponded to the 50th percentile of the estimated logistic curve. We computed confidence intervals (CIs) to determine the experimental conditions in which the difference between crossover values for A1 and A2 differed reliably from zero.

## RESULTS

We examined whether speech sounds modify the perception of sensations arising in the facial skin. To address this issue, we carried out a sensory test in which a robotic device gently stretched the facial skin lateral to the oral angle in combination with the presentation of speech sounds. Figure 2 shows the effects of auditory inputs on the perception of two skin stretch stimuli of equal magnitude (1 N in each case in *experiment 1*). The ordinate in Fig. 2 shows the mean normalized probability of judging the second stretch as greater. As shown in Fig. 2, compared with skin stretch judgments in the absence of auditory speech inputs (CTL), the mean probability ( $\pm$ SE) of judging the skin stretch magnitude as greater increased when the stretch occurred in conjunction with the word “had” (A1) and decreased when the second stretch accompanied the word “head” (A2). ANOVA indicated that the probability of judging the second stretch as greater differed across conditions [ $F_{(2,14)} = 13.17, P < 0.001$ ]. Post hoc statistical comparisons found reliable differences in probability between A1 and A2 ( $P < 0.001$ ) and between A2 and CTL ( $P < 0.05$ ). The difference in probability between A1 and CTL was not statistically reliable ( $P = 0.085$ ). Speech sounds thus appear to bias

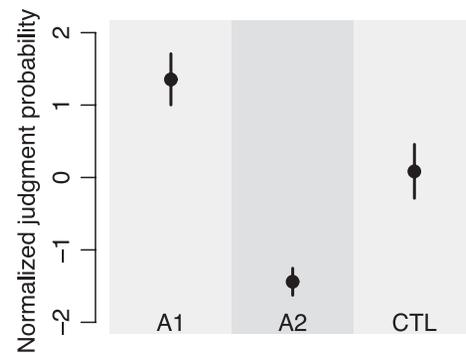


Fig. 2. Speech sounds change the perception of facial skin sensation. The graph shows the mean normalized probability of judging the second stretch as greater. The *left* point shows the probability when the second stretch was paired with the sound “had” (A1). The *middle* point shows the probability when the second stretch was paired with the sound “head” (A2). The *right* point shows stretch magnitude judgments in the absence of auditory input [control (CTL)]. Error bars show SEs across the participants.

somatosensory perception. Subjects felt more force when the skin stretch accompanied the sound “had” than the sound “head.”

*Experiment 2* examined whether the link that we observed between auditory and somatosensory information is specifically tied to auditory speech inputs and speechlike patterns of facial skin deformation or whether instead the interaction lacks in specificity. Figure 3A shows representative examples of perceptual judgments for individual participants in the four experimental conditions. The circles in Fig. 3A show the probability of judging the second skin stretch as greater for different magnitudes of stretch. The solid symbols shown the judgment probabilities when the sound “had” is presented with the second skin stretch (A1), and the shaded symbols show the probabilities when the sound “head” is presented with the second stretch (A2; Fig. 3A). The solid and shaded lines showed the estimated psychometric functions in the two audio conditions, respectively (Fig. 3A). In all four experimental conditions, participants’ judgment probabilities gradually changed from 0 to 1 depending on the force difference between the skin stretch stimuli. Thus, in all cases, participants were able to successfully perform the force judgment task. The *top left* graph in Fig. 3A shows that speech sounds affect the perception of somatosensory inputs arising in facial skin. Skin stretch that occurred in conjunction with the sound “had” was consistently judged as greater than skin stretch that occurred in conjunction with the sound “head.” In contrast, the use of acoustically matched but nonspeech stimuli resulted in a smaller somatosensory effect. Judgments of skin stretch applied to the arm or the palm were not affected at all by simultaneous speech input.

Psychometric functions fit to the mean data in the A1 and A2 audio conditions were assessed quantitatively using logistic regression. A reliable difference between the two functions was observed in the face-speech condition [ $\chi^2(1) = 7.06, P < 0.01$ ]. The psychometric functions were not reliably different in the face-nonspeech [ $\chi^2(1) = 2.73, P = 0.098$ ], forearm-speech [ $\chi^2(1) = 1.10, P > 0.25$ ], or palm-speech [ $\chi^2(1) = 0.12, P > 0.9$ ] conditions.

The effects of auditory input and skin stretch were also assessed in terms of differences in the mean crossover force value (the force difference at which the two skin stretch stimuli

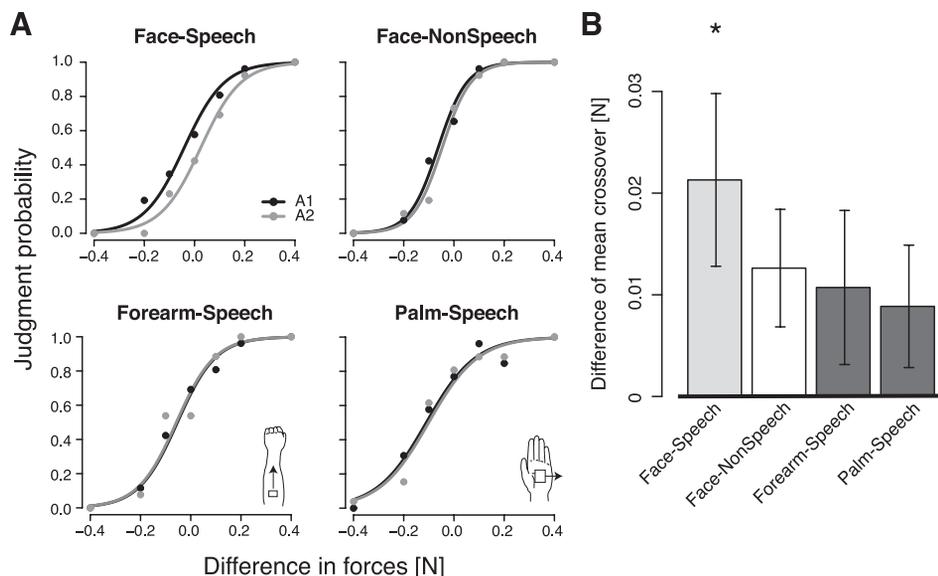


Fig. 3. Changes in the perception of skin stretch are specific to speech sounds and facial skin sensation. *A*: example of changes to the perception of skin stretch due to auditory speech stimuli. The abscissa shows the difference in magnitude of the two skin stretch stimuli. The ordinate shows the mean probability of judging the second stimulus as greater. Solid circles indicate the judgments in the A1 (“had-head”) condition. Shaded circles indicate judgments in the A2 (“had-head”) condition. The solid and shaded lines show the estimated psychometric functions. *B*: difference in mean identification crossover. Error bars show SEs across the participants. \*Conditions in which auditory stimuli reliably change the probability of judging the second stretch as greater.

were judged to be equal in magnitude; Fig. 3*B*). For purposes of this analysis, we assessed on a per subject basis the crossover, or 50% point of the psychometric function that was fit to the probability of judging the second stretch as greater, as a function of the force difference between the first and second stretch (Fig. 3*B*). Somatosensory judgments for the facial skin were reliably affected by speech sounds. The mean difference in crossover value between A1 and A2 was 0.021 N ( $P < 0.01$ , 99% CI: 0.0001 and 0.042 N). As above, pure tones had a marginal effect on the crossover value for somatosensory judgments ( $P = 0.08$ , mean value: 0.013 N, 99% CI:  $-0.008$  and 0.033 N). There was no reliable effect of speech sounds on somatosensory judgments involving the skin of the forearm ( $P > 0.10$ , mean value: 0.011 N, 99% CI:  $-0.010$  and 0.032 N) or the palm ( $P > 0.10$ , mean value: 0.009 N, 99% CI:  $-0.013$  and 0.03 N).

In summary, we found that the perception of facial skin stretch was modified by speech sounds. When the skin was stretched in synchrony with the presentation of the word “had,” subjects judged that the skin stretch force was greater than when the skin stretch occurred in conjunction with word “head.” The somatosensory response was reliably altered by speech sounds and was less affected by matching nonspeech auditory inputs. No somatosensory modulation was observed when the same speech stimuli accompanied skin stretch applied to the forearm or palm.

## DISCUSSION

The central finding of this study was the demonstration that the perception of facial skin deformation that typically occurs in conjunction with speech production is modified by speech sounds. The modulation is specific to facial skin sensation and also speech sound perception. The finding should be viewed in conjunction with a complementary demonstration that facial somatosensory inputs alter the perception of speech sounds (Ito et al. 2009). The two findings taken together suggest that sound and proprioception are integrated in the neural processing of speech. Facial somatosensory inputs affect the way that speech signals sound, and auditory inputs alter that way that speech movements feel. The effect of auditory inputs on the somato-

sensory system is selective. Speech sounds may possibly serve to tune the motor system in the acquisition process of speech production.

We found that speech sounds altered judgments of facial skin stretch while the effects of nonspeech sounds on skin stretch judgment were marginal. This latter finding could be consistent with the possibility that sound characteristics affect somatosensory function at a relatively early stage of processing. However, filters in biological systems are never abrupt. Hence, the finding is also consistent with the idea that nonspeech sounds that are speechlike in nature, as is the case here, are processed by the same neural mechanisms that deal with speech. Presumably as the difference between speech and nonspeech stimuli increases, the relative effects on orofacial somatosensory processing would be expected to decline.

We considered the possibility that the effects observed here are attentional rather than perceptual in nature. This seems to be unlikely. The subjects’ task is to make a somatosensory judgment based on stimulation delivered to the face, forearm, or palm. The very same auditory input (speech) is applied in all three conditions. The attentional requirements are similar for all three tasks. The subject is required to judge the somatosensory stimulus magnitude in the presence of speech. We found that the same speech stimuli had an effect on somatosensory perceptual judgments when stimulation was delivered to the face but not when skin on the palm or forearm was stretched. This suggests that the effects are perceptual in nature rather than due to attentional differences between the conditions.

There have been previous demonstrations of auditory-somatosensory interactions in situations other than speech (Jousmaki and Hari 1998; Ro et al. 2009; Yau et al. 2009; Zampini and Spence 2004). In particular, there are robust effects of auditory input on tactile frequency discrimination, and the auditory cross-sensory interference is quite specific. Audio distracters strongly impair frequency discrimination of vibrotactile stimulation to the finger, but only if the frequencies of the auditory and tactile stimuli are similar. Audio distracters do not interfere with judgments of tactile intensity (Yau et al. 2009). The specificity of the cross-talk between the auditory and somatosensory systems is consistent with the present

finding that somatosensory effects were dependent on the specific interaction between speech sounds and somatosensory judgments related to the facial skin. The word “had,” which is associated with larger jaw opening movements than the word “head,” led to reports of greater skin stretch, suggesting that the modulation of somatosensory perception is related to the articulatory motion that is associated with the sound that was heard. These various examples of intersensory interactions point to the importance of supramodal representations for somatosensory-auditory interactions that are specifically tuned to the context of the integration.

Interactions between auditory and somatosensory information may well be relevant to the neural processing of speech since speech processes and certainly speech production involves both auditory information and inputs that arise from the muscles and tissues of the vocal tract. Recent theoretical models of speech processing have assumed neural linkages between the speech perception and production systems (Guenther et al. 2006; Hickok et al. 2009). However, studies of this relationship have focused largely on the influence of motor function on auditory speech processing and, more specifically, on the interaction of auditory inputs and speech motor cortex (D’Ausilio et al. 2009; Fadiga et al. 2002; Meister et al. 2007; Watkins et al. 2003; Wilson et al. 2004; Yuen et al. 2009). The possible involvement of the somatosensory system in speech perceptual processing has largely been ignored. The present results are complementary with those of our previous demonstration that somatosensory inputs affect the perception of speech sounds (Ito et al. 2009). The two sets of results in combination suggest that at a neural level, speech-relevant somatosensory processing lies at the intersection between speech production and perception. Indeed, the data are consistent with the possibility that the somatosensory system forms part of the neural substrate of both processes. Moreover, these effects occur in the absence of articulatory motion or speech production (cf. Champoux et al. 2011). The findings thus suggest a new role for the somatosensory system in speech perception that is largely independent of motor function.

The brain areas involved in the auditory-somatosensory interaction that is observed here are not known. The neural substrate may be in belt areas of auditory association cortex, notably the caudomedial area, based on demonstrations of audio-tactile interactions in humans (Foxe et al. 2002; Murray et al. 2005) and nonhuman primates (Fu et al. 2003; Kayser et al. 2005). The second somatosensory cortex may also be responsible for the observed sensory modulation as it is activated by skin stretch (Backlund Wasling et al. 2008) and modulated by auditory inputs (Lütkenhöner et al. 2002). Cortical regions such as the planum temporale (Hickok et al. 2009) and subcortical structures such as the superior colliculus (Stein and Meredith 1993) may also be involved in somatosensory-auditory interaction. Studies to date of the neural substrates of auditory-somatosensory interactions have focused on brain regions involved in nonspeech processes. Further investigation is required to know the extent to which these regions are involved in the current speech-relevant somatosensory-auditory interaction.

#### GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R03-DC-009064 and R01-DC-04669.

#### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

#### AUTHOR CONTRIBUTIONS

Author contributions: T.I. and D.J.O. conception and design of research; T.I. performed experiments; T.I. analyzed data; T.I. and D.J.O. interpreted results of experiments; T.I. prepared figures; T.I. and D.J.O. drafted manuscript; T.I. and D.J.O. edited and revised manuscript; T.I. and D.J.O. approved final version of manuscript.

#### REFERENCES

- Backlund Wasling H, Lundblad L, Löken L, Wessberg J, Wiklund K, Norrsell U, Olausson H. Cortical processing of lateral skin stretch stimulation in humans. *Exp Brain Res* 190: 117–124, 2008.
- Beauchamp MS, Yasar NE, Frye RE, Ro T. Touch, sound and vision in human superior temporal sulcus. *Neuroimage* 41: 1011–1020, 2008.
- Champoux F, Shiller DM, Zatorre RJ. Feel what you say: an auditory effect on somatosensory perception. *PLoS One* 6: e22829, 2011.
- D’Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. *Curr Biol* 19: 381–385, 2009.
- Diehl RL, Lotto AJ, Holt LL. Speech perception. *Annu Rev Psychol* 55: 149–179, 2004.
- Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci* 15: 399–402, 2002.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol* 88: 540–543, 2002.
- Fu KMG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE. Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci* 23: 7510–7515, 2003.
- Gick B, Derrick D. Aero-tactile integration in speech perception. *Nature* 462: 502–504, 2009.
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96: 280–301, 2006.
- Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101: 2725–2732, 2009.
- Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69: 407–422, 2011.
- Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley, 2000.
- Ito T, Tiede M, Ostry DJ. Somatosensory function in speech perception. *Proc Natl Acad Sci USA* 106: 1245–1248, 2009.
- Jousmäki V, Hari R. Parchment-skin illusion: sound-biased touch. *Curr Biol* 8: R190, 1998.
- Kayser C, Petkov CI, Augath M, Logothetis NK. Integration of touch and sound in auditory cortex. *Neuron* 48: 373–384, 2005.
- Lakatos P, Chen C, O’Connell MN, Mills A, Schroeder CE. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53: 279–292, 2007.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev* 74: 431–461, 1967.
- Lütkenhöner B, Lammertmann C, Simões C, Hari R. Magnetoencephalographic correlates of audiotactile interaction. *Neuroimage* 15: 509–522, 2002.
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 264: 746–748, 1976.
- Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. *Curr Biol* 17: 1692–1696, 2007.
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15: 963–974, 2005.
- Ro T, Hsu J, Yasar NE, Elmore LC, Beauchamp MS. Sound enhances touch perception. *Exp Brain Res* 195: 135–143, 2009.
- Remez RE, Rubín PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science* 212: 947–949, 1981.

- Stein BE, Meredith MA.** *The Merging of the Senses*. Cambridge, MA: MIT Press, 1993.
- Sumby WH, Pollack I.** Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26: 212–215, 1954.
- Tajadura-Jiménez A, Kitagawa N, Väljamäe A, Zampini M, Murray MM, Spence C.** Auditory-somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra. *Neuropsychologia* 47: 195–203, 2009.
- Watkins KE, Strafella AP, Paus T.** Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41: 989–994, 2003.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M.** Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7: 701–702, 2004.
- Yau JM, Olenczak JB, Dammann JF, Bensmaia SJ.** Temporal frequency channels are linked across audition and touch. *Curr Biol* 19: 561–566, 2009.
- Yau JM, Weber A, Bensmaia S.** Separate mechanisms for audio-tactile pitch and loudness interactions. *Front Psychol* 1: 2010.
- Yuen I, Davis MH, Brysbaert M, Rastle K.** Activation of articulatory information in speech perception. *Proc Natl Acad Sci USA* 107: 592–597, 2010.
- Zampini M, Spence C.** The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *J Sensor Stud* 19: 347–63, 2004.

