Somatosensory function in speech perception

Takayuki Ito[a], Mark Tiede[a,b], and David J. Ostry[a,c,1]

[a]Haskins Laboratories, 300 George Street, New Haven, CT 06511; [b]Research Laboratory of Electronics, Massachusetts Institute of Technology, 50 Vassar Street, Cambridge, MA 02139; and [c]Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, Canada H3A 1B1

Edited by Fernando Nottebohm, The Rockefeller University, Millbrook, NY, and approved December 4, 2008 (received for review October 7, 2008)

Somatosensory signals from the facial skin and muscles of the vocal tract provide a rich source of sensory input in speech production. We show here that the somatosensory system is also involved in the perception of speech. We use a robotic device to create patterns of facial skin deformation that would normally accompany speech production. We find that when we stretch the facial skin while people listen to words, it alters the sounds they hear. The systematic perceptual variation we observe in conjunction with speech-like patterns of skin stretch indicates that somatosensory inputs affect the neural processing of speech sounds and shows the involvement of the somatosensory system in the perceptual processing in speech.

multisensory integration | speech production

Somatosensory information arising from facial skin deformation is a significant but little-recognized source of sensory input accompanying speech production. The present study indicates that there is also somatosensory involvement in the perception of speech. The existence of pathways between somatosensory and auditory areas of the brain has been documented previously (1–5). However, little is known about the possible role of somatosensory inputs, and in particular, those associated with speech production, in the perception of speech sounds.

The effects of somatosensory inputs on auditory function have been reported previously in phenomena unrelated to speech. For example, vibrotactile inputs to the hands affect judgments of perceived loudness (5). There are likewise multisensory influences on auditory perception in speech. The contribution of visual inputs in particular is well documented (6, 7). However, possible somatosensory effects on speech perception are different in that somatosensory inputs are not involved in any obvious way in the perception of speech sounds.

Somatosensory feedback in speech production is intriguing because of the lack of muscle proprioceptors in orofacial muscles (8–11). In the absence of the usual pattern of sensory support from muscle proprioceptors, other somatosensory inputs are presumably important to orofacial motor control. Recently there has been growing recognition of the importance of information associated with skin deformation in limb motor control (12–14). Because the facial skin is rich in cutaneous mechanoreceptors (15) and is systematically deformed during normal speech production (16), somatosensory input associated with skin deformation may be important in articulatory control, and possibly in speech perception as well.

The idea that perception and production are mediated by common mechanisms originates in the motor theory of speech perception (17, 18). However, the possible effects of somatosensory input on speech perception are almost entirely unknown, and indeed evidence to date for the link between perception and production comes exclusively from demonstrations of cortical motor activation in conjunction with speech perception (19–21). The potential effect of somatosensory input on perception would be a good indication that inputs normally associated with production are also involved in speech perceptual processing.

In the present study, we show that speech-like patterns of facial skin deformation affect the way people hear speech sounds. Our subjects listened to words one at a time that were taken from a computer-generated continuum between the words *head* and *had*. We found that perception of these speech sounds varied in a systematic manner depending on the direction of skin stretch. The presence of any perceptual change at all depended on the temporal pattern of the stretch, such that perceptual change was present only when the timing of skin stretch was comparable to that which occurs during speech production. The findings are consistent with the hypothesis that the somatosensory system is involved in the perceptual processing of speech. The findings underscore the idea that there is a broad nonauditory basis to speech perception (17–21).

Results

We examined whether we could change the perceptual boundary between 2 words by using speech-patterned somatosensory input that was produced by stretching the facial skin as subjects listened to the stimulus words. We used a robotic device (Fig. 1) to create patterns of facial skin deformation that would be similar to those involved in the production of *head* and *had*. We tested 3 directions of skin stretch (up, down, backward) with 3 different groups of subjects. We also carried out 2 control tests in which we assessed the effects on speech perception of patterns of skin stretch that would not be experienced during normal facial motion in speech. One involved a twitch-like pattern of skin stretch; the other involved static skin deformation.

We found that the perceptual boundary between *head* and *had* was altered by the skin stretch perturbation. Fig. 2 shows, for a single subject, a representative example of the perceptual modulation using a single cycle of 3-Hz skin stretch in an upward direction. The figure shows the probability that the subject judged the stimulus word as *had* for each of the 10 stimulus levels on the *head* to *had* continuum. The fitted lines give the estimated psychometric function, with the control condition shown in blue and the skin stretch condition shown in red. As can be seen, the value corresponding to the 50th percentile of the estimated psychometric function shifted to the right in the perturbation condition. This means that the upward skin stretch resulted in an increase in the probability that a word was identified as *head*. It can also be seen that the skin stretch perturbation had different effects at different points along the stimulus continuum. The perceptual effect of the skin stretch was greatest for intermediate values of *head* and *had* and had little effect near the ends of the continuum, where the stimuli were clearly identified as one word or the other.

We found that speech sound classification changed in a predictable way depending on the direction in which we stretched the skin. When the skin was stretched upward, the stimulus was more often judged as *head*. When the skin was stretched downward, the stimulus sounded more like *had*. When the stretch was in a backward direction, there was no perceptual

Author contributions: T.I., M.T., and D.J.O. designed research; T.I. performed research; T.I. contributed new reagents/analytic tools; T.I. and D.J.O. analyzed data; and T.I., M.T., and D.J.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: david.ostry@mcgill.ca

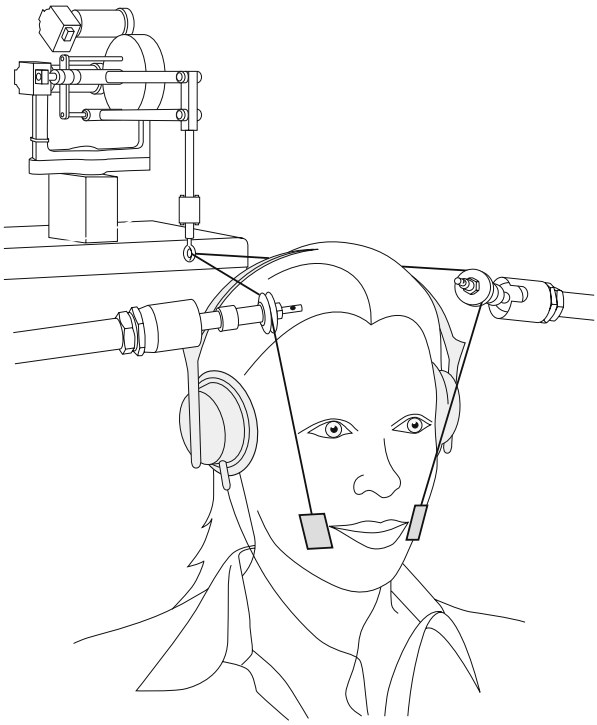

Fig. 1. Experimental setup for the delivery of skin stretch perturbations.

effect. Fig. 3 summarizes the mean difference between the control and the perturbation conditions for the 3 directions of 3-Hz skin stretch by using mean probability (Fig. 3*A*) and identification cross-over (Fig. 3*B*) as dependent measures. The error bars give the standard error of the difference between means (i.e., between perturbation and control conditions). We assessed perceptual differences due to skin stretch and skin stretch direction by using a 2-way ANOVA, with one repeated factor (perturbation versus control trials) and one between the subject's factor (skin stretch direction). Bonferonni-corrected comparisons showed that when we stretched the facial skin upward, the identification cross-over value increased and the mean probability of responding *had* decreased ($P < 0.01$ in both cases). When we stretched the facial skin downward, we observed the opposite pattern. Specifically, the identification cross-

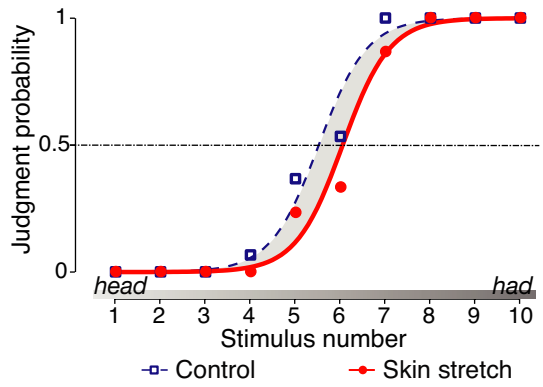

Fig. 2. Representative changes to speech perception with a 3-Hz sinusoidal pattern of upward skin stretch. The blue squares are for judgments without skin stretch. The red circles show judgments that occur in conjunction with skin stretch. Squares and circles show the judgment probability for each audio stimulus. The two solid lines show the estimated psychometric functions.

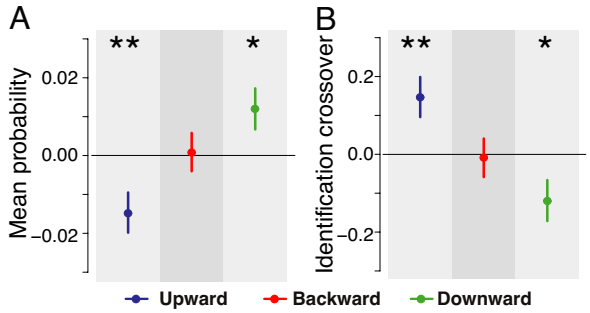

Fig. 3. Perceptual classification of speech sounds depends on the direction of skin stretch. The graphs show differences in mean probability (*A*) and category boundary (identification cross-over frequency) (*B*) with and without skin stretch. Error bars show the standard error across subjects. The asterisks indicate significant differences between control and skin stretch conditions (**, $P < 0.01$; *, $P < 0.05$).

over value decreased and the mean probability of responding *had* increased ($P < 0.025$ for both post hoc tests). Finally, when we stretched the skin backward there was no effect on identification performance ($P > 0.85$). These systematic changes in speech sound classification as a function of skin stretch direction indicate that somatosensory information associated with the facial skin stretch can play a role in the perception of speech sounds. The modulation suggests a close tie between the neural processes of speech perception and production.

We assessed the extent to which the perceptual classification of speech sounds was dependent on the specific temporal pattern of facial skin stretch. For the data described above, we used 3-Hz sinusoidal patterns of skin stretch to approximate the temporal pattern of facial skin deformation that would normally occur in conjunction with the jaw lowering and raising movements in the production of *head* and *had*. Here we examined whether the skin stretch pattern had to be speech-like if changes in auditory classification were to be observed. We compared the previously obtained responses by using 3-Hz skin stretch with 2 new patterns: a single cycle of 9-Hz stretch and a pattern involving static skin stretch (see *Methods*). We focused on the effects of skin stretch in an upward direction because this had previously produced the largest perceptual effects.

Fig. 4 summarizes the difference between control and perturbation conditions in terms of mean probability (Fig. 4*A*) and identification cross-over value (Fig. 4*B*). The error bars show the

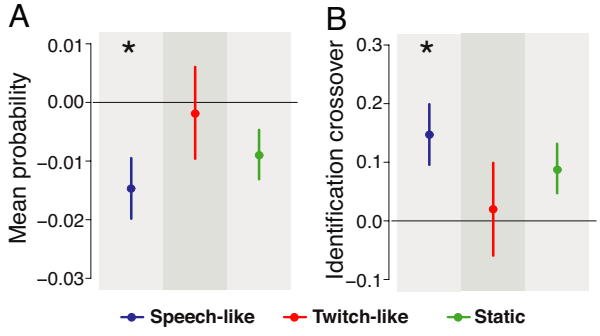

Fig. 4. Changes in speech perception are dependent on a speech-like time course of facial skin deformation. The graphs show differences in mean probability (*A*) and category boundary (identification cross-over frequency) (*B*) between stretch and no-stretch conditions for 3 different patterns of facial skin deformation. Error bars give the standard error across subjects. Asterisks indicate significant differences between stretch and no-stretch conditions. (*, $P < 0.05$).

standard error of the difference between means across subjects. Data for the 3-Hz condition are repeated from the previous section. Two-way ANOVA followed by Bonferroni-corrected comparisons was used to test for differences between skin stretch and control conditions for the 3 temporal patterns of stretch. In contrast to the reliable perceptual modulation observed in response to the 3-Hz pattern ($P < 0.02$), there was no perceptual change associated with either the 9-Hz twitch-like stretch pattern ($P > 0.75$) or the static stretch pattern ($P > 0.10$). This result suggests that perceptual modulation of speech sounds by somatosensory information is dependent on a speech-like temporal pattern of somatosensory change. To the extent that this can be extended to speech production and perception more generally, it would suggest that the effects of production on speech perception are movement specific.

Discussion

The principal finding of the present study is that the perception of speech sounds is modified by stretching the facial skin and that the perceptual change depends on the specific pattern of deformation. Moreover, the presence of any perceptual change depended on the temporal pattern of the stretch such that perceptual change was present only when the timing of skin stretch was comparable to that which occurs during speech production.

Evidence to date for the idea that speech production and perception systems have a common neural substrate has come entirely from work on motor function and has been motivated by the motor theory of speech perception (17, 18). For example, studies using transcranial magnetic stimulation (TMS) have recently documented the involvement of the premotor and motor cortex in the neural processing of speech (19–21). These studies have shown that the evoked electromyographic response to TMS to the lip area of motor cortex is facilitated by watching speech movements and listening to speech sounds (20), and that repetitive TMS to premotor cortex affects performance in a speech perception task (21). In contrast, the possible effects of somatosensory function on speech perception have been unexplored despite extensive evidence that somatosensory inputs are important in speech production as well as in motor function more generally. Present findings show that speech perception is linked not only to cortical motor areas involved in speech production but importantly is also affected by the kinds of somatosensory inputs that would normally arise in conjunction with speech production.

The modulation of speech perception observed in the present study may arise as a consequence of facial somatosensory input to facial motor and premotor areas, a pattern that would be consistent with the motor theory of speech perception (17, 18). However, somatosensory inputs may also affect auditory processing more directly. A number of studies have reported bidirectional effects linking somatosensory and auditory cortices (2–5). Indeed, activity due to somatosensory inputs has been observed within a region of the human auditory cortex that is traditionally considered unisensory (3). It is also possible that the observed perceptual effects arise at the subcortical level, in particular, in the superior colliculus, which is considered a site of multisensory integration, including auditory-somatosensory interaction (1).

The role of somatosensory inputs in speech perception was investigated by using a different approach in the facial somatosensory system (i.e., by stretching the facial skin). Somatosensory function associated with facial skin deformation is a little-recognized source of orofacial kinesthesia. However, the facial skin is rich in cutaneous mechanoreceptors (15) and is systematically deformed in the context of normal speech production (16). Somatosensory inputs associated with facial skin deformation are a primary source of sensory support for speech motor function owing to the lack of muscle proprioceptors in many orofacial muscles (8–11).

The neural responses of cutaneous mechanoreceptors are dependent on the direction of sensory input both in the orofacial system and in the hands (22). Moreover, in recent studies of somatosensory function in the limbs (12–14) it has been reported that stretching the skin results in a sensation of movement if the skin is stretched in a manner corresponding to normal movement. Different patterns of skin stretch at the side of the mouth may similarly bias the perception in the associated direction. Stretching the skin lateral to the oral angle, which is the area that we focused on in the present study, induced a cortical reflex that was associated with a modification of lip position in response to a sudden change in the position of the jaw (23). Because the cutaneous mechanoreceptors lateral to the oral angle are activated during speech movements and especially jaw motion (15), stretching the facial skin lateral to the oral angle could provide kinesthetic information concerning articulatory motion. This information may both complement and shift the perception of the speech sounds.

Methods

Seventy-five native speakers of American English participated in the experiment. The subjects were all young adults, had normal hearing, and had no neurological deficits. There were 5 separate experimental conditions, and 15 different subjects were tested in each condition. All subjects signed the approved Yale University Human Investigation Committee informed consent form.

Auditory Stimuli. The stimulus continuum was generated by using an iterative Burg algorithm for estimating spectral parameters (24). The procedure involved shifting the first (F1) and the second (F2) formant frequencies in equal steps from values observed for *head* to those associated with *had*. The stimuli were generated from tokens provided by a male native speaker of English. For this individual, we obtained average values across 5 tokens of *head* of 537 Hz and 1640 Hz, for F1 and F2, respectively. Values for *had* were 685 Hz and 1500 Hz, respectively.

Speech Perception Test. On each trial, the subject was presented with one of the 10 words, selected in random order. The task was to identify whether the word was *head* or *had* by pressing a button on a display screen with a computer mouse. The probability that the subject answered *had* was calculated for each of the 10 stimuli that formed the continuum and the obtained proportions were fitted with a logistic function (25). A screening test that used the same stimuli was conducted before the main experiment. The purpose was to verify that judgment probabilities changed monotonically over the stimulus set. Subjects that failed to display a monotonic psychometric function were excluded from the main test. Three subjects were eliminated on this basis.

Skin Stretch Perturbation. We programmed a small robotic device (Phantom 1.0, SensAble Technologies) to apply skin stretch loads (Fig. 1). The skin stretch was produced by using small plastic tabs (2×3 cm), which were attached bilaterally to the skin at the sides of the mouth and were connected to the robotic device through thin wires. The wires were supported by wire supports with pulleys to avoid contact between the wires and the facial skin. By changing the configuration of the robotic device and the wire supports, the facial skin was stretched in different directions.

We have focused on sensory inputs arising in the facial skin because there are systematic facial skin deformations in conjunction with speech production (16). The facial skin is rich with cutaneous afferents, but their role as a source of sensory information in speech production is infrequently recognized. We have focused specifically on the skin at the sides of the mouth for a variety of reasons. Infraorbital nerve afferents with cutaneous receptive fields at this location are activated by speech production (15). Cutaneous afferents at the side of the mouth are also implicated in jaw movement. In particular, cutaneous mechanoreceptors at the oral angle have been shown to respond to passive jaw motion (26, 27). Skin stretch at this location results in a compensatory reflex response that is normally evoked by unpredictable jaw position change (23).

The temporal pattern and timing of the facial skin deformation were as follows. The temporal profile was that of a single cycle of a 3-Hz sinusoid that was chosen to approximate the duration of a jaw opening–closing cycle in

speech. The commanded amplitude was 4 N. This resulted in 10–15 mm of skin stretch. The timing of the facial skin stretch relative to the audio signal was set, on the basis of preliminary testing, in order that subjects perceived both auditory and somatosensory signals simultaneously. A difference of 90 ms in start time (with the skin stretch perturbation first) was used in all tests involving sinusoidal force patterns.

We carried out two control tests: one involving a twitch-like skin stretch, the other involving static skin deformation. In the case of the twitch-like skin stretch, the profile was a single cycle of a 9-Hz sinusoidal pattern. The amplitude and peak timing of the facial skin stretch were the same as in the 3-Hz condition. The 9-Hz stretch resulted in a perturbation that was shorter than the sounds of our auditory stimuli and faster than normal speech movements. In the case of static skin deformation, a constant force of 2 N deformed the facial skin while the subject listened to the stimulus word. The resulting skin stretch amplitude was comparable to that used with sinusoidal loads, about 10–15 mm. The force onset was 730 ms before the audio onset. The force was removed when the subject pressed the answer button. Both the twitch force and the constant force were applied in an upward direction.

Experimental Procedure. In each of the conditions of force application, we alternated blocks of perturbation trials (10 trials) with blocks of control trials (10 trials). In the perturbation trials, perceptual judgments were accompanied by facial skin stretch. In control trials, the subject was required to perform the same perceptual judgments but no loads were applied. Each subject was tested with one direction of skin stretch and one temporal pattern of force. In total there were 600 trials (and thus 600 perceptual judgments) per experiment, 300 in the control (no skin stretch) condition and 300 in the skin stretch condition. Thirty responses were recorded for each of the 10 steps between *head* and *had*.

Statistical Analysis. We carried out quantitative tests for differences in perceptual performance between perturbation and control conditions by using two different measures: (*i*) mean judgment probability and (*ii*) the identification cross-over value. The mean judgment probability was calculated by averaging the judgment probabilities of all stimulus levels in each condition; the identification cross-over (category boundary) was determined by finding the auditory stimulus value corresponding to the 50th percentile of the estimated psychometric function. We evaluated differences between control and perturbed conditions by using a repeated-measures ANOVA in which control trials versus perturbed trials were within the subject's factors and the skin stretch direction or the temporal pattern of skin stretch were between the subject's factors. ANOVA was followed by Bonferonni-corrected comparisons to assess the reliability of specific pairwise differences.

ACKNOWLEDGMENTS. We thank Douglas N. Honorof, Rafael Laboissière, Andrew A. G. Matter, David W. Purcell, and Bruno Repp for advice and assistance. This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-04669.

1. Stein BE, Meredith MA (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
2. Foxe JJ, *et al.* (2002) Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *J Neurophysiol* 88:540–543.
3. Murray MM, *et al.* (2005) Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15:963–974.
4. Jousmaki V, Hari R (1998) Parchment-skin illusion: Sound-biased touch. *Curr Biol* 8:R190.
5. Schurmann M, Caetano G, Jousmaki V, Hari R (2004) Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. *J Acoust Soc Am* 115:830–832.
6. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748.
7. Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215.
8. Folkins JW, Larson CR (1978) In search of a tonic vibration reflex in the human lip. *Brain Res* 151:409–412.
9. Neilson PD, Andrews G, Guitar BE, Quinn PT (1979) Tonic stretch reflexes in lip, tongue and jaw muscles. *Brain Res* 178:311–327.
10. Stål P, Eriksson PO, Eriksson A, Thornell LE (1987) Enzyme-histochemical differences in fibre-type between the human major and minor zygomatic and the first dorsal interosseus muscles. *Arch Oral Biol* 32:833–841.
11. Stål P, Eriksson PO, Eriksson A, Thornell LE (1990) Enzyme-histochemical and morphological characteristics of muscle fibre types in the human buccinator and orbicularis oris. *Arch Oral Biol* 35:449–458.
12. Edin BB, Johansson N (1995) Skin strain patterns provide kinaesthetic information to the human central nervous system. *J Physiol* 487:243–251.
13. Collins DF, Prochazka A (1996) Movement illusions evoked by ensemble cutaneous input from the dorsum of the human hand. *J Physiol* 496:857–871.
14. Collins DF, Refshauge KM, Todd G, Gandevia SC (2005) Cutaneous receptors contribute to kinesthesia at the index finger, elbow, and knee. *J Neurophysiol* 94:1699–1706.
15. Johansson RS, Trulsson M, Olsson KA, Abbs JH (1988) Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Exp Brain Res* 72:209–214.
16. Connor NP, Abbs JH (1998) Movement-related skin strain associated with goal-oriented lip actions. *Exp Brain Res* 123:235–241.
17. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
18. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36.
19. Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur J Neurosci* 15:399–402.
20. Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41:989–994.
21. Meister IG, *et al.* (2007) The essential role of premotor cortex in speech perception. *Curr Biol* 17:1692–1696.
22. Edin BB, Essick GK, Trulsson M, Olsson KA (1995) Receptor encoding of moving tactile stimuli in humans. I. Temporal pattern of discharge of individual low-threshold mechanoreceptors. *J Neurosci* 15:830–847.
23. Ito T, Gomi H (2007) Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *Neuroreport* 18:907–910.
24. Orfanidis SJ (1988) *Optimum Signal Processing, an Introduction* (Macmillan, New York).
25. Dobson AJ (1990) *An Introduction to Generalized Linear Models* (CRC, Boca Raton, FL).
26. Appenteng K, Lund JP, Seguin JJ (1982) Behavior of cutaneous mechanoreceptors recorded in mandibular division of Gasserian ganglion of the rabbit during movements of lower jaw. *J Neurophysiol* 47:151–166.
27. Ro JY, Capra NF (1995) Encoding of jaw movements by central trigeminal neurons with cutaneous receptive fields. *Exp Brain Res* 104:363–375.