

Integration of auditory and somatosensory error signals in the neural control of speech movements

Yongqiang Feng, Vincent L. Gracco and Ludo Max

J Neurophysiol 106:667-679, 2011. First published 11 May 2011; doi:10.1152/jn.00638.2010

You might find this additional info useful...

This article cites 40 articles, 10 of which can be accessed free at:

<http://jn.physiology.org/content/106/2/667.full.html#ref-list-1>

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

<http://jn.physiology.org/content/106/2/667.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 September 5, 2011.

Integration of auditory and somatosensory error signals in the neural control of speech movements

Yongqiang Feng,¹ Vincent L. Gracco,^{2,3,5} and Ludo Max^{4,5}

¹Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China; ²Centre for Research on Brain, Language, and Music and ³School of Communication Sciences and Disorders, McGill University, Montreal, Quebec, Canada; ⁴Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington; and ⁵Haskins Laboratories, New Haven, Connecticut

Submitted 19 July 2010; accepted in final form 11 May 2011

Feng Y, Gracco VL, Max L. Integration of auditory and somatosensory error signals in the neural control of speech movements. *J Neurophysiol* 106: 667–679, 2011. First published May 11, 2011; doi:10.1152/jn.00638.2010.—We investigated auditory and somatosensory feedback contributions to the neural control of speech. In *task I*, sensorimotor adaptation was studied by perturbing one of these sensory modalities or both modalities simultaneously. The first formant (F1) frequency in the auditory feedback was shifted up by a real-time processor and/or the extent of jaw opening was increased or decreased with a force field applied by a robotic device. All eight subjects lowered F1 to compensate for the up-shifted F1 in the feedback signal regardless of whether or not the jaw was perturbed. Adaptive changes in subjects' acoustic output resulted from adjustments in articulatory movements of the jaw or tongue. Adaptation in jaw opening extent in response to the mechanical perturbation occurred only when no auditory feedback perturbation was applied or when the direction of adaptation to the force was compatible with the direction of adaptation to a simultaneous acoustic perturbation. In *tasks II* and *III*, subjects' auditory and somatosensory precision and accuracy were estimated. Correlation analyses showed that the relationships 1) between F1 adaptation extent and auditory acuity for F1 and 2) between jaw position adaptation extent and somatosensory acuity for jaw position were weak and statistically not significant. Taken together, the combined findings from this work suggest that, in speech production, sensorimotor adaptation updates the underlying control mechanisms in such a way that the planning of vowel-related articulatory movements takes into account a complex integration of error signals from previous trials but likely with a dominant role for the auditory modality.

sensorimotor integration; adaptation; speech motor control; auditory feedback; somatosensory feedback

SENSORY INFORMATION plays a critical role in learning motor skills, adapting to new environments, and maintaining accurate performance of motor activities such as reaching, grasping, and speaking. Empirically, it has been demonstrated that both limb and speech motor behavior are affected by perturbed sensory feedback. One approach has examined changes in motor behavior as a result of unexpected alterations of sensory feedback or perturbations of the effectors. For speech, such studies have observed immediate compensatory adjustments in the perturbed as well as nonperturbed articulators in response to mechanical perturbations of the lips or jaw (Folkens and Abbs 1975; Abbs and Gracco 1984), an altered hard palate (Honda et

al. 2002), and altered auditory feedback (Burnett et al. 1998; Purcell and Munhall 2006b). A different approach involves perturbing sensory feedback or effector motion in a consistent rather than unexpected manner and examining the process of sensorimotor adaptation, a gradual learning of adjusted movements that achieve the desired sensory consequences. Adapted responses typically are predictive (observed from the beginning of the movement) rather than reactive and are characterized by after-effects when normal feedback is restored (Cohen 1966; Shadmehr and Mussa-Ivaldi 1994). Thus, the central nervous system takes the altered environment into account during movement planning and appears to learn a new mapping between sensory consequences and motor commands.

For speech production, sensorimotor adaptation has been demonstrated with auditory feedback manipulations (Houde and Jordan 1998, 2002; Max et al. 2003; Shiller et al. 2009), structural changes of the vocal tract (Baum and McFarland 1997; Jones and Munhall 2003; Hamlet et al. 1976; Savariaux et al. 1995), and mechanical perturbations of jaw movement trajectory (Nasir and Ostry 2006; Tremblay et al. 2003, 2008). After similar work with whispered speech (Houde and Jordan 1998, 2002), studies in which the formant frequencies (vowel-specific vocal tract resonances) in the real-time auditory feedback signal were digitally shifted up or down during normally voiced speech showed that subjects adjusted their articulation such that the produced formant frequencies were modified in the opposite direction of the experimental manipulation (Max et al. 2003). Adaptation and after-effects occurred during the production of isolated vowels as well as meaningful words. Moreover, adaptation was measurable at vowel onset, occurred with both suddenly and incrementally introduced perturbations, occurred after only a few trials, differed in extent across different vowels, and decreased in extent for very large perturbations. Other researchers have replicated selected aspects of these findings in studies in which only the first formant (F1) was shifted (Purcell and Munhall 2006a; Villacorta et al. 2007). Interestingly, data from one study suggest that subjects who show more extensive adaptation may have better auditory acuity (Villacorta et al. 2007).

For force field perturbations that affect jaw motion during speech without measurable effects on the associated acoustic output, the reported data were obtained after hundreds of learning trials, and even then 30% or more of the subjects showed no adaptation (Nasir and Ostry 2006; Tremblay et al. 2003). In fact, Nasir and Ostry (2006) noted that “about 70% showed improvement (or at least no deterioration) in perfor-

Address for reprint requests and other correspondence: L. Max, Dept. of Speech and Hearing Sciences, Univ. of Washington, 1417 N.E. 42nd St., Seattle, WA 98105-6246 (e-mail: LudoMax@uw.edu).

mance with training. The remainder got worse with training.” Such findings may be consistent with the long-standing hypothesis that, at least for vowel production, the achievement of task goals is monitored mainly in the auditory domain (Gracco and Abbs 1986; Guenther et al. 1999, 2006; Perkell et al. 1993, 2000, 2008). It is noteworthy in this context that, to date, formant shifts in the feedback signal have been demonstrated to lead to adaptation despite the fact that this requires atypical movement patterns, whereas adaptation to jaw force fields has been tested only when neither the perturbation nor the adaptation resulted in measurable changes of the speech acoustics. Thus, although directly related to fundamental questions regarding the relative weighting of auditory versus somatosensory (specifically kinesthetic) feedback signals in the neural control of speech movements, it remains unknown whether force field perturbations would still result in similar adaptation if adapting could only be accomplished at the cost of atypical speech acoustics.

Here, we investigated this question of multimodality sensory integration in speech production directly by examining speech acoustics and kinematics in conditions with only an auditory perturbation, only a somatosensory perturbation, simultaneous compatible auditory and somatosensory perturbations (i.e., feedback in the two modalities provides nonconflicting information about the type of error), or simultaneous incompatible auditory and somatosensory perturbations (i.e., feedback in the two modalities provides conflicting information about the type of error). Auditory perturbations consisted of upward shifts in F1 frequency (a higher F1 frequency normally results from a greater extent of oral opening), and somatosensory perturbations consisted of loads applied to the jaw in the direction of either opening or closing. Uncompensated loads in the opening direction would lead to more oral opening, and thus these loads would further increase the amount of auditory error. As a result, adapting to such a downward force on the jaw would also assist with reducing auditory error. Uncompensated loads in the closing direction would reduce oral opening, and thus these loads would decrease the amount of auditory error. As a result, adapting to an upward force on the jaw would oppose the reduction of auditory error, whereas not adapting to the force would reduce auditory error. The primary experimental question was whether adaptation to a simultaneous manipulation of F1 and jaw position serves to minimize mainly the auditory error or somatosensory error. A secondary goal was to examine the potential relationship, if any, between auditory or somatosensory acuity and sensorimotor adaptation.

MATERIALS AND METHODS

Subjects

Subjects were eight adults (4 women and 4 men) from 18 to 31 yr of age (mean: 22 yr, SD: 4.4 yr). All subjects were native speakers of American English who reported no speech, language, hearing, or neurological disorders and who did not take medications that affect sensorimotor functioning. Behavioral pure tone hearing thresholds were 20 dB hearing level or lower for the frequencies of 250 Hz to 8 kHz. Subjects completed a pretest and *task I* (a task with four sensorimotor adaptation conditions) on 1 day and *task II* (an auditory acuity task) and *task III* (a somatosensory acuity task) 10–14 days later. All subjects gave informed consent before participation, and all study procedures were approved by the local Institutional Review Board.

Pretest and Task I

Pretest procedures. A pretest was conducted with each subject to estimate the acoustic and kinematic parameters needed for the implementation of subsequent *tasks I* and *III* and to examine the effect on speech acoustics and kinematics when jaw perturbations were delivered unpredictably (i.e., without sensorimotor adaptation). Subjects’ speech was recorded with a microphone (WL185, Shure), amplifier (DI/O Preamp system 267, ART), and a high-quality soundcard (HDSP 9632, RME) installed in a personal computer. Jaw movements were recorded and perturbed with a robotic device (Phantom Premium 1.0A, SensAble Technologies). The distal tip of the robot arm was attached to the jaw by means of an aluminum rotary coupler and a mandibular dental appliance individually made for each subject. A separate maxillary dental appliance, held in place by two adjustable mechanical arms, was used in conjunction with a head support system to immobilize the head.

First, each subject was asked to briefly close the jaw and then to read aloud a short passage. The closed position and trajectory of the jaw movements were recorded by the robot’s optical encoders, and principal component analysis was used to compute the major axis of movement. All recordings of jaw position in the robot’s three-dimensional coordinate system were transformed to single-dimension values by projecting to this major axis and using the jaw closed position as the origin. Second, subjects completed two productions of each of the words “bat,” “hack,” and “pap” to provide data for estimating the individual’s maximum jaw opening (O_m). Third, subjects produced eight trials of a set of four words with the vowels /ε/ or /æ/ (“bet,” “heck,” “bat,” and “hack”) while an unpredictable jaw perturbation was applied during some of the trials. There were 16 trials without perturbation, 8 trials with an upward force, and 8 trials with a downward force.

Each trial started when the subject moved the jaw to a zone of acceptable start positions (a 0.4-mm range centered around a position 1 mm below the closed position) represented by a horizontal bar on a computer monitor. Visual feedback about jaw position was provided in the form of a horizontal line that had to be moved into the bar. When the jaw was held in the starting zone for 1 s, a word appeared and remained visible for 2 s. Subjects were instructed to say the word during this time interval. The monitor then turned blank for 0.3 s before the visual indicator of the start position appeared again for the next trial.

During word production, jaw perturbations were implemented by the Phantom robot as a displacement-dependent force (see Fig. 1) for which each individual subject’s O_m determined the stiffness ($k = 1/O_m$ N/mm). Forces were applied to the jaw along the principal axis of movement with the strength of the force proportional to the distance between the instantaneous jaw displacement and a threshold such that $f = ak|p - T|$, if $P < T$; and $f = 0$, if $P > T$, where f is force, a is direction (+1 for up and -1 for down), k is the rate of change in force magnitude over displacement (i.e., stiffness), p is jaw displacement relative to the start position, and T is the threshold below which the force is nonzero. T was 1 mm below the actual start position. The robot’s control software was programmed to limit the maximum force magnitude to 1 N for all subjects.

Task I procedures. In *task I*, a sensorimotor adaptation task, subjects read aloud monosyllabic consonant-vowel-consonant words while auditory feedback, somatosensory feedback, or both parameters were experimentally manipulated during the perturbation phase of each condition. F1 in the auditory feedback was either not manipulated or shifted up (i.e., causing auditory feedback to indicate increased oral opening), and the robot applied to the jaw no force, an upward force (i.e., in the direction of closing), or a downward force (i.e., in the direction of opening). Given that not all combinations of these manipulations were necessary to address the hypotheses under investigation, combined with the need to minimize total time of subject discomfort associated with head immobilization and the in-

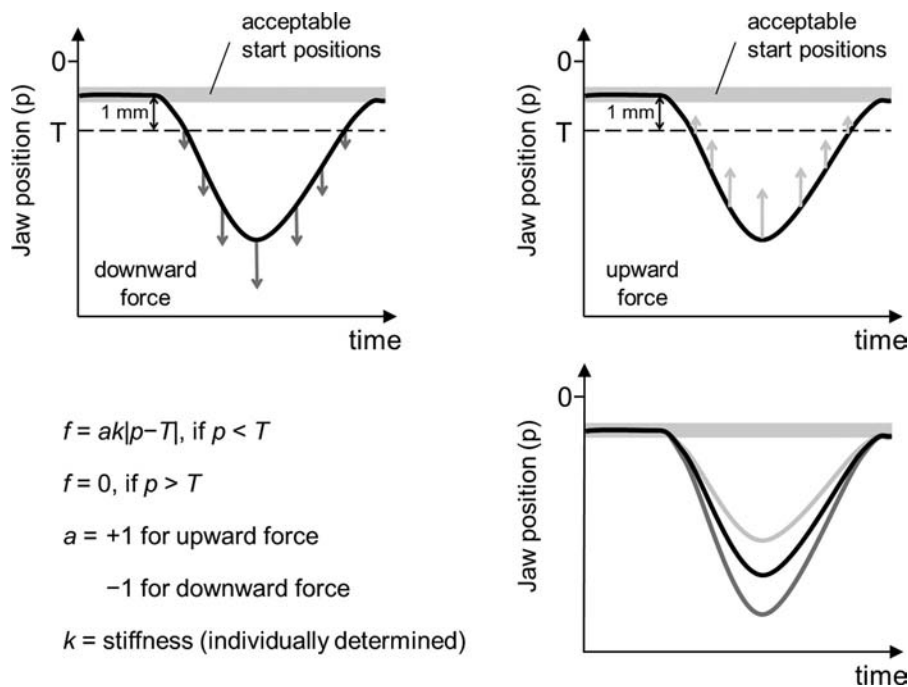


Fig. 1. *Top*: stylized illustrations of downward (*left*, dark gray arrows) or upward (*right*, light gray arrows) force (f) as a function of the distance between jaw position (p) and threshold (T) located 1 mm below the position at movement onset. The shaded area indicates the range of acceptable start positions (each trial's stimulus word was presented when the jaw was held within this range for 1 s). *Bottom right*: schematic illustration of the resulting changes in jaw movement extent in the presence of downward (dark gray trace) or upward (light gray trace) forces relative to a nonperturbed condition (black trace).

sertion of appliances and sensors in the oral cavity, four 15-min experimental conditions were included in randomized order: F1 no shift with a downward force, F1 up shift with an upward force, F1 up shift with a downward force, and F1 up shift with no force.

To examine adaptation, training words with the vowels /*e*/ and /*æ*/ (“bet,” “heck,” “bat,” and “hack”) were produced repeatedly in pre-perturbation (baseline), perturbation, and post-perturbation (after-effects) phases. Stimulus presentation and jaw movement start positions were controlled as described above for the pretest. Within each condition, there were 30 epochs of training words (cycles in which each of the 4 words was produced once in randomized order): 4 epochs in the baseline phase, 15 epochs while the magnitude of the perturbation(s) was incrementally increased across trials (ramp perturbation phase), 7 epochs while the perturbation was delivered at its maximum value (max perturbation phase), and 4 epochs in the post-perturbation phase. To examine the generalization of after-effects, test words with different vowels or consonants (“hawk,” “hick,” “pap,” and “pep”) were produced once before and once after the perturbation phase. Test words were inserted in the last two epochs of the baseline phase and first two epochs in the post-perturbation phase. Across the two adjacent epochs, four training words and four test words were presented in random order.

F1 in the auditory feedback was raised by routing the acoustic signal through filters and a signal processor before being fed back to the subject. The second (F2) and higher formants as well as the fundamental frequency and consonant-related noise components remained unaltered. Specifically, the signal was routed through parallel low-pass (LP) and high-pass (HP) filters with a common cutoff frequency (f_c), defined as: $f_c = \sqrt{F1F2}$. Filtering was implemented by two dual-channel units (model 852, Wavetek) so that the cascaded channels of one unit yielded the LP band and the cascaded channels of the other unit yielded the HP band. For each subject, f_c was set individually based on F1 and F2 measures for unperturbed trials from the pretest. The average of the subject's f_c values for /*e*/ and /*æ*/ was then used as the common f_c value for the LP and HP filters.

The LP filter output was sent to a digital processor (VoiceOne, TC Helicon) capable of shifting the formants up in real time (measured delay: 10 ms) under computer control. The amount of upward shift in the max perturbation phase was 400 cents (100 cents = 1 semitone) after a gradual increase from 0 cents in steps of 10 cents during the

ramp phase. The output from the processor and the unaltered output from the HP filter were mixed (32–1100A Stereo Mixer, Realistic) to recreate a full spectrum. This signal was routed to insert earphones (ER-3A, Etymotic Research) via a headphone amplifier (S-phone, Samson) and also recorded as one channel on a notebook computer via an external audio interface (model US-144, Tascam) at a sampling rate of 44.1 kHz. In addition to this altered acoustic signal, an unaltered output from the microphone amplifier was recorded simultaneously on a second channel of the audio interface and notebook computer.

Intensity of the sound output in the earphones was calibrated before each experimental session. First, the energy ratio of the LP and HP bands of the reconstructed acoustic signal was adjusted to exactly match that of the unaltered microphone amplifier output. The gain of the total signal pathway was then adjusted so that an input of 75 dB sound pressure level [SPL(A)] at the microphone (15 cm from the subject's mouth) resulted in output of 72 dB SPL(A) in the insert earphones as measured in a 2-cc coupler connected to a sound level meter (model 2700, Quest Technologies).

Jaw movements associated with vowel production in the training words were perturbed by applying upward or downward forces with the robot in the same manner as described above for the pretest with the exception that jaw load forces were now incrementally increased during the ramp phases. Each individual subject's k value for the max perturbation phase was determined as previously described, and k was incrementally increased from 0 to this maximum k . In postexperiment interviews, all subjects reported that they had been unaware of either the auditory or somatosensory perturbations.

Data collection and processing. KINEMATIC MEASURES. An electromagnetic midsagittal articulograph (model AG200, Carstens) was used to transduce jaw, lip, and tongue movements. One sensor was attached 1 cm from the tip of the tongue (T1), and two sensors were placed more posteriorly along the midline with a spacing of ~15 mm (T2 and T3). Sensors were also attached to the lower dental appliance (jaw movement) and to the upper and lower lips. Sensors serving as fixed reference points were secured to the upper dental appliance and to the bridge of the nose. Cyanodent was used to attach sensors to the tongue and dental appliances, and double-sided adhesive tape was used to attach sensors to the lips and bridge of the nose. At the end of

each recording session, the maxillary occlusal plane was determined with two sensors attached to a bite-plate.

The x,y coordinates of reference sensors were LP filtered at 5 Hz, and those of all other sensors at 15 Hz. Movement onset and offset, respectively, were automatically identified as the time points where a sensor's velocity exceeded 10 mm/s and where movement direction reversed, as indicated by a local minimum in the velocity trace. Jaw height for the vowel was defined as coordinate y of the jaw sensor in the maxillary occlusal plane-based coordinate system at the moment of jaw opening movement offset (maximum opening or peak displacement). At the same time point, we also measured the y -coordinate of the three sensors on the tongue. Finally, for separate analyses, motion of the three sensors on the tongue was also mathematically decoupled from the jaw (Westbury et al. 2002) to allow separate analyses of possible adaptive changes in tongue positioning that occurred independent from the jaw. Before statistical analysis, all kinematic measures for each condition were normalized by subtracting their corresponding mean value in the baseline phase. For every epoch, the normalized measures were then averaged across the two words with the same vowel.

ACOUSTIC MEASURES. Acoustic recordings were first downsampled to 10 kHz. The first two formants (F1 and F2) were automatically extracted by custom Matlab software (The Mathworks) but with visual inspection and manual correction if necessary. The onset and offset of each vowel were labeled based on visual inspection of a spectrogram. An order-optimized linear predictive coding analysis (Vallabha and Tuller 2002) was conducted on a 30-ms Hamming window centered at time points 40% and 50% into the vowel for each trial. Formant frequency estimates from the two windows were averaged to obtain the dependent variables (F1 and F2) for that trial. These acoustic measures were normalized in cents relative to the mean frequency in the baseline phase, and, for every epoch, the normalized measures were then averaged across the two words with the same vowel.

Statistical analyses. Data from the pretest with unpredictable jaw perturbations were analyzed to verify that loads of the magnitude used here would actually affect the extent of jaw opening, and possibly also F1, in the absence of adaptation. Therefore, paired t -tests were used to examine the changes in jaw position or F1 for the no perturbation versus upward perturbation and no perturbation versus downward perturbation conditions.

For *task I*, the primary question was under which combinations of auditory and somatosensory perturbations acoustic and/or kinematic adaptation would occur. Given that each subject's baseline mean value for each dependent variable was used as the reference to normalize the data (i.e., each subject's baseline mean value for each dependent variable was always 0), paired t -tests comparing baseline and max perturbation data effectively corresponded to one-sample t -tests evaluating whether or not the max perturbation data deviated significantly from 0.

It should be noted that the auditory perturbation altered the auditory feedback signal without directly altering the acoustics or kinematics of the subjects' productions. Therefore, adaptation to this perturbation would be reflected in statistically significant changes from the baseline value in the acoustic and/or kinematic recordings (adjustments that would yield more typical feedback). Similarly, after-effects caused by the auditory perturbations would be reflected in a continued use of the adaptive response (different from baseline) in the postperturbation phase. In contrast, the jaw forces applied for the somatosensory perturbation would directly alter the subject's speech kinematics if left uncompensated. Hence, adaptation to this jaw perturbation would be reflected in an absence of statistically significant changes from baseline in the kinematic-dependent variables (indicating that the central nervous system compensated for the force and maintained relatively typical productions and associated somatosensory feedback).

A significance level of 0.05 was selected a priori to balance the probability of type I and type II errors. Additionally, Cohen's d (Cohen 1988) was calculated as a measure of effect size. For the one-sample case, $d = m/\sigma$, where m is the mean and σ is the SD, respectively. Small, medium, and large effect sizes correspond to d values equal to or larger than 0.2, 0.5, and 0.8, respectively.

Task II

Data collection, processing, and statistical analysis. *Task II* investigated whether or not there is a relationship between subjects' extent of acoustic adaptation in response to F1 shifts in the auditory feedback and their auditory acuity for F1. Auditory acuity was examined using the method of adjustment (Gescheider 1976): for each trial, the subject was presented with a stimulus pair that contained a standard stimulus and a comparison stimulus, and the subject was asked to adjust the comparison stimulus until it was perceived to be identical to the standard stimulus. The comparison stimuli differed from the standard stimulus in F1.

Two representative trials (one word with / ε / and one word with / α /) from the subject's unperturbed productions in *task I* were chosen as standard stimuli. The standard stimulus was always played back from the personal computer sound card (RME HDSP 9632) and routed through the same parallel filters, vocal processor (not implementing an F1 shift), mixer, headphone amplifier, and earphones as used in *task I*. At the beginning of each trial, the subject heard this standard stimulus and was asked to memorize it for subsequent comparison. A comparison stimulus was then generated by playing back the same production through the same instrumentation but with implementation of an F1 shift by the processor. The amount of shift for a given trial was randomly selected from the set of $-400, -390, -380, \dots, 390, 400$ cents. The subject was instructed to compare the comparison stimulus with the standard stimulus in memory and, if any discrepancy was perceived, to adjust the comparison stimulus by pressing the buttons of a joystick to increase or decrease the comparison stimulus F1. Subjects were allowed to relisten to the comparison stimulus and adjust it until they thought that it matched the standard stimulus. The final F1 value of the comparison stimulus was recorded. For each of the two vowels, 50 pairs were presented. The order of blocked testing for / ε / and / α / was randomized for each subject.

The mean (\bar{x}) and SD (σ) of F1 for a subject's adjusted comparison stimuli (in cents relative to the standard stimulus) were calculated for each vowel. In psychophysical terms, \bar{x} is the point of subjective equality (PSE) and σ is the difference limen (DL). PSE and the derived constant error (CE; the difference between PSE and the standard stimulus) represent auditory accuracy (bias in average perception), whereas DL represents auditory precision (spread in perception around an average value). Here, CE has the same value as PSE because F1 of the standard stimulus is 0 cents.

In one experimental condition of *task I*, F1 in the auditory feedback was raised without simultaneous jaw perturbation. Thus, correlation analyses (Pearson's r for / ε / and / α / separately) were conducted to determine if the extent of acoustic adaptation in the F1 up no force condition of *task I* was related to auditory precision (DL) or accuracy (CE) as measured in *task II*. Given that CE values of equal magnitude but opposite sign indicate equally large misjudgments, individual subjects' absolute values were used for the CE variable (note that DL is always positive).

Task III

Data collection, processing, and statistical analysis. *Task III* was designed to investigate whether or not adaptation to a jaw perturbation is related to somatosensory acuity for jaw position. Standard and comparison jaw positions were presented to the subject through passive movements implemented by the Phantom robot, and the

subject was asked to match the standard jaw position by actively adjusting the jaw comparison position.

For each subject, the lower (J_{-1}) and upper (J_{+1}) limits of the tested range corresponded to the average jaw opening for productions with a downward or upward perturbation, respectively, in the pretest. The average jaw opening for unperturbed productions (J_0) was used as the standard stimulus that was presented on each trial. Comparison stimuli were jaw positions ranging from J_{-1} to J_{+1} . For each of 50 trials per vowel (/ɛ/ and /æ/ were tested separately with their own J_{-1} , J_0 , and J_{+1}), a random value for the comparison jaw position was generated by the Matlab-based experiment control software.

At the beginning of each trial, the robot moved the subject's jaw passively from the closed position to the standard position. The subject clicked a joystick button when this position had been memorized. The robot then moved the jaw back to the closed position and, after another joystick click, from there to the comparison position. The subject was allowed to take as much time as desired to adjust this position by actively moving the jaw and then clicked the button to indicate that this was perceived as identical to the standard position. The robot implemented all passive movements with a duration of 200 ms. There were 50 trials for each "vowel" (no sound was produced), and the order of testing the "vowels" was again randomized.

Presentation of the standard jaw position was influenced by resistance from the subject, and, thus, the standard position was not always identical. Given that each comparison position was adjusted against the standard position presented in the same individual trial, such trial-dependent variation was removed by measuring the difference between the standard and adjusted comparison positions. The mean and SD of the jaw position difference were calculated to represent CE

and DL, respectively. Given that one experimental condition in *task I* involved downward jaw perturbations without simultaneous auditory perturbation, correlation analyses (Pearson's r for each vowel) were used to determine if the extent of kinematic adaptation in the jaw during that experimental condition in *task I* was related to the subject's somatosensory precision (DL) or accuracy (CE) as tested in *task III*. Analogous with *task II*, individual subjects' absolute values were used for the CE variable.

RESULTS

Pretest

Effects of the applied force on the extent of jaw opening, and possibly on F1, in the absence of sensorimotor adaptation were examined during the pretest. Figure 2, *top*, shows kinematic data (jaw position relative to the closed position) for each individual subject during words with the vowels /ɛ/ (*left*) and /æ/ (*right*) in the presence of a downward force, no force, or upward force. Lowered jaw positions in the presence of a downward force and higher jaw positions in the presence of an upward force were observed for all eight subjects. Consequently, for both vowels, statistically significant differences were found between jaw position with an upward or downward force compared with no force (Table 1). Figure 2, *bottom*, shows the corresponding individual subject acoustic data (F1). Neither direction of force caused F1 in the perturbed trials to be statistically significantly different from the unperturbed trials,

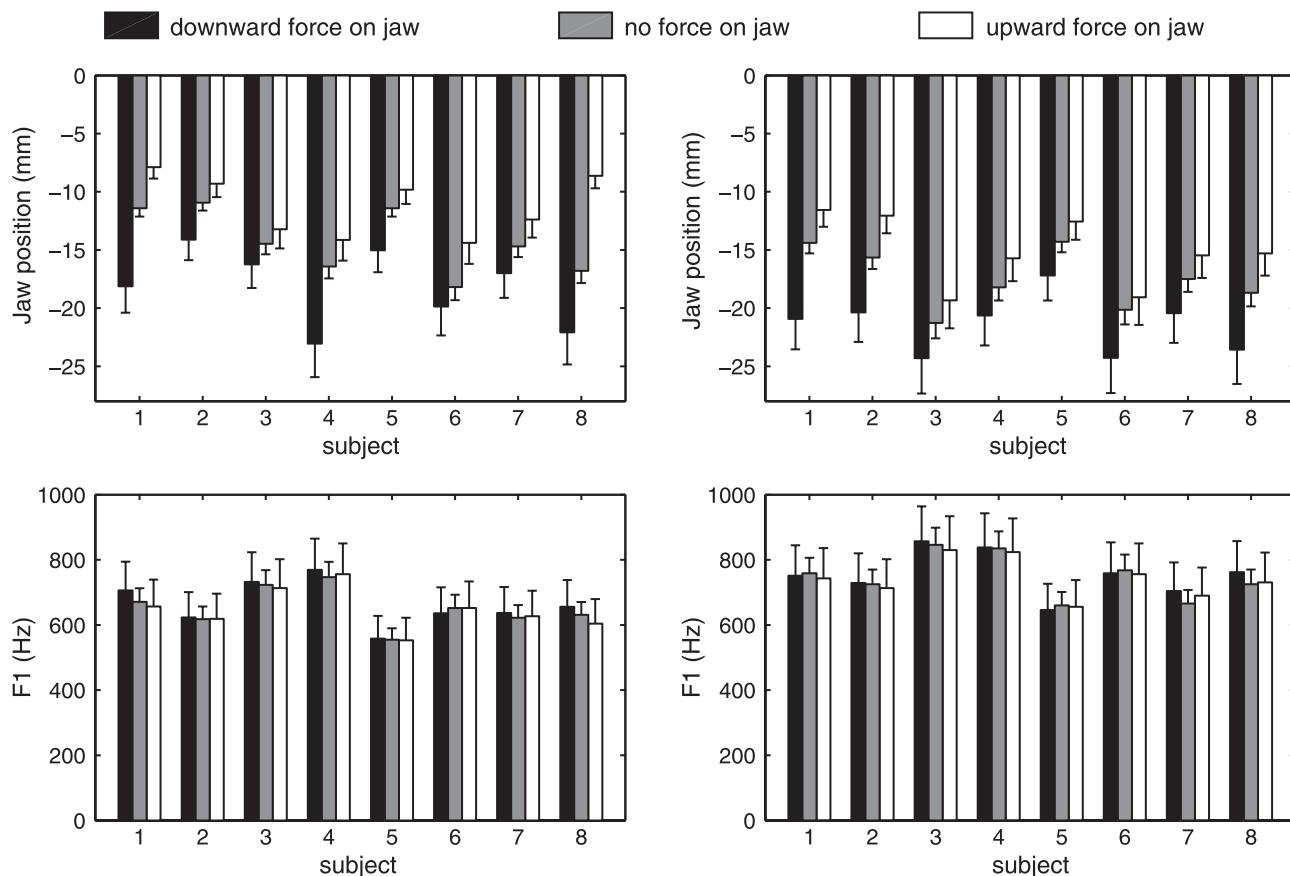


Fig. 2. Individual subject data for jaw position at peak displacement (*top*) and first formant (F1) frequency (*bottom*) during /ɛ/ in "bet" and "heck" (*left*) and /æ/ in "bat" and "hack" (*right*) with and without unpredictable force perturbations on the jaw. Solid, shaded, and open bars indicate productions with a downward force perturbation, without perturbation, and with an upward force perturbation, respectively. Error bars indicate SEs. Subjects 1–4 were women; subjects 5–8 were men.

Table 1. *t* and *P* values of paired *t*-tests and *d* values for jaw displacement with an unpredictable downward or upward perturbation force versus without force in the pretest

	/ɛ/		/æ/	
	Force down versus no force	No force versus force up	Force down versus no force	No force versus force up
<i>t</i> (df = 7)	-5.337	-3.855	-8.014	-7.894
<i>P</i>	0.001	0.006	<0.001	<0.001
<i>d</i>	1.9	1.4	2.8	2.8

df, degrees of freedom; *d*, effect size.

most likely as a result of long-latency reflex-based compensation in nonperturbed articulators (Abbs and Gracco 1984; Ito et al. 2005).

Sensorimotor Adaptation in Acoustic Dependent Variables

F1 frequencies for /ɛ/ and /æ/ that have been normalized and averaged across consonant contexts and subjects are shown in Fig. 3. Excluding the test words (produced to examine generalization rather than adaptation), there were 30 epochs: epochs 1–4 in the baseline phase, epochs 5–19 in the ramp phase, epochs 20–26 in the max perturbation phase, and epochs 27–30 in the postperturbation phase. In all three F1 up conditions, all eight subjects lowered F1 in their productions during the ramp and max perturbation phases and then showed after-effects in the form of a gradual (rather than sudden) return to near baseline values during the postperturbation phase. That is, adaptive responses were temporarily continued even after normal feedback had been restored. Across these three conditions and across all subjects and both vowels (3 conditions × 8 subjects × 2 vowels = 48 adaptation measures), the mean amount of adaptation in F1 during the max perturbation phase (during which the processor-applied F1 shift was 400 cents)

Table 2. *t* and *P* values of paired *t*-tests and *d* values for normalized F1 frequencies of /ɛ/ and /æ/ in the baseline phase versus the max perturbation phase

	F1 Up, No Force	F1 Up, Force Up	F1 Up, Force Down	F1 No Shift, Force Down
<i>/ɛ/</i>				
<i>t</i> (df = 7)	-6.463	-4.357	-4.283	1.715
<i>P</i>	<0.001	0.003	0.004	0.130
<i>d</i>	2.3	1.5	1.5	0.6
<i>/æ/</i>				
<i>t</i> (df = 7)	-8.774	-3.524	-8.583	1.470
<i>P</i>	<0.001	0.010	<0.001	0.185
<i>d</i>	3.1	1.3	3.0	0.5

F1, first formant.

was -110.26 cents (SD: 59.51). The results shown in Fig. 3 suggest that the extent of adaptation might have increased further if subject comfort considerations had not prevented the inclusion of more trials.

Table 2 shows the results and associated effect sizes for the *t*-tests examining for which conditions the normalized F1 frequencies in the max perturbation phase deviated in a statistically significant manner from their baseline value. As mentioned above, a statistically significant change from baseline in an acoustic variable would indicate that adaptation did occur for the condition under investigation. As shown in Table 2, the F1 data from the max perturbation phase yielded such statistically significant results (i.e., adaptation) for all three conditions in which F1 in the auditory feedback was experimentally shifted up.

Adaptive changes in F2 [note that F2 was not manipulated but was evaluated because it has been suggested to covary with F1 so as to maintain a vowel-specific formant ratio (see Miller 1989)] were also analyzed after normalizing and averaging this dependent variable in the same manner as described above for F1. Statistically significant increases in F2 occurred in the F1

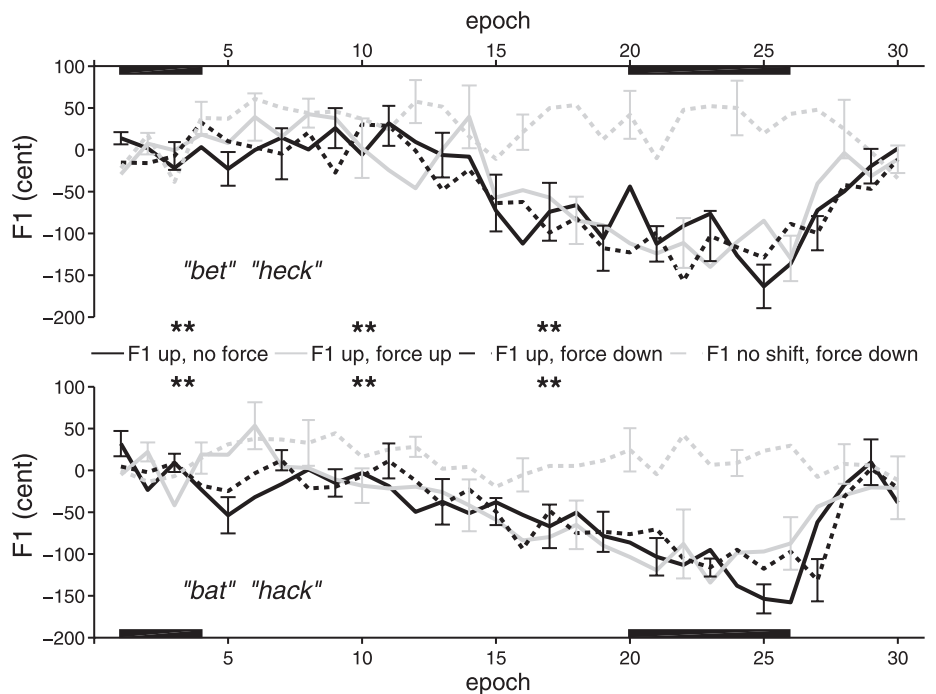


Fig. 3. Group mean F1 frequency for /ɛ/ (“bet” and “heck”) and /æ/ (“bat” and “hack”) in four conditions. Individual subject data were normalized relative to the baseline mean value using the latter as the reference when converting frequency measures from Hertz to cents. Error bars indicate SE. ***P* < 0.05 and Cohen’s *d* > 0.5 (see Table 2).

Table 3. *t* and *P* values of paired *t*-tests and *d* values for normalized F2 frequencies of / ϵ / and / \ae / in the baseline phase versus the max perturbation phase

	F1 Up, No Force	F1 Up, Force Up	F1 Up, Force Down	F1 No Shift, Force Down
<i>/ϵ/</i>				
<i>t</i> (df = 7)	3.274	4.128	2.241	1.143
<i>P</i>	0.014	0.004	0.060	0.290
<i>d</i>	1.2	1.5	0.8	0.4
<i>/\ae/</i>				
<i>t</i> (df = 7)	2.038	3.819	0.972	-0.500
<i>P</i>	0.081	0.007	0.363	0.632
<i>d</i>	0.7	1.4	0.3	0.2

F2, second formant.

up no force condition and in the F1 up force up condition for / ϵ / as well as in the F1 up force up condition for / \ae / (Table 3). Descriptively, these F2 increases were small in light of the F1 decreases reported above: across conditions, subjects, and vowels, the mean F2 increase was 32.02 cents (SD: 36.38).

Generalization in Acoustic Dependent Variables

F1 for the test words produced in the F1 up no force condition (the only condition without force perturbation) was analyzed to examine whether or not subjects showed generalization of acoustic after-effects to other vowels and consonant contexts. Note that generalization of after-effects would be reflected in a continuation of the adaptive use of lowered F1 frequencies for words that were produced after normal feedback had already been restored for the training words and that, unlike the training words, were never produced with F1-shifted feedback. As shown in Fig. 4, group mean F1 values were indeed lower in the postperturbation phase than in the baseline phase for all four test words, resulting in negative normalized values postperturbation. With the present sample size, this decrease in F1 reached or approached statistical significance for two of the four test words, namely, those test words that differed from two training words in vowel context (Table 4).

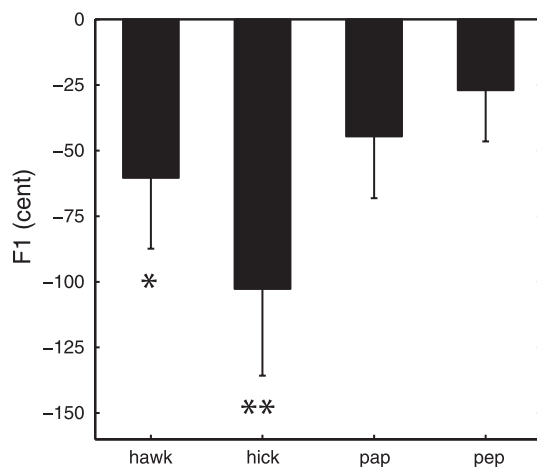


Fig. 4. Group mean F1 frequency for four test words in the postperturbation phase normalized to baseline mean value. Test words were never produced with formant-shifted feedback and served to examine generalization. Error bars indicate SEs. ***P* < 0.05; **P* < 0.07.

Table 4. *t* and *P* values of paired *t*-tests and *d* values for normalized F1 frequencies of test words included to examine generalization of acoustic after-effects in the postperturbation phase versus the baseline phase

	Test Words			
	“hawk”	“hick”	“pap”	“pep”
<i>t</i> (df = 7)	-2.231	-3.106	-1.890	-1.379
<i>P</i>	0.061	0.017	0.101	0.217
<i>d</i>	0.8	1.1	0.7	0.5

Sensorimotor Adaptation in Kinematic Dependent Variables

Normalized jaw positions at the time of peak displacement for / ϵ / and / \ae / are shown in Fig. 5. In the F1 up force up condition, this jaw position measurement increased gradually as the upward force increased during the ramp phase, and jaw position in the max perturbation phase was statistically significantly elevated above the baseline value (Table 5). In other words, jaw position was significantly altered by the perturbing force. A descriptive comparison with the extent of jaw position changes caused by unpredictable perturbations in the pretest, as shown in Fig. 2, revealed that there was indeed little compensation for the perturbing force in this F1 up force up condition; thus, minimal adaptation to the force took place. In contrast, there were no statistically significant changes in jaw position during the max perturbation phase versus the baseline phase for any of the remaining three conditions even though a downward force was applied to the jaw in two of those conditions (thus, adaptation to the force did take place in those two conditions such that the typical jaw position was maintained despite the applied load).

Position data for the anterior tongue sensor (T1) during / ϵ / and / \ae / are shown in Fig. 6, and the corresponding statistics are shown in Table 6. When a downward force was applied to the jaw, the T1 vertical position was not lowered from baseline to max perturbation, regardless of whether or not F1 was shifted up. However, the same T1 position was raised in the condition where F1 was shifted up and no force was applied. In addition, the same T1 vertical position was also raised in a similar manner when F1 was shifted up and an upward force was applied to the jaw. The middle (T2) and posterior (T3) tongue sensors showed similar increases in height during this F1 up force up condition.

As shown in Table 7, when the T1 data were mathematically decoupled from the jaw (representing independent tongue movements that increased or decreased the amount of movement caused by the tongue’s mechanical coupling to the jaw), there was a strong trend for the T1 position to increase when F1 was shifted up and no force was applied to the jaw (a condition without statistically significant change in jaw position). This trend for the F1 up no force condition was observed for both vowels. However, there were no such trends when F1 was shifted up and an upward force was simultaneously applied to the jaw (the condition with a statistically significant increase in jaw position) or in any of the other conditions. Statistical analyses completed on the decoupled data for T2 and T3 showed neither significant changes nor trends toward a change in the position of these sensors with respect to the jaw.

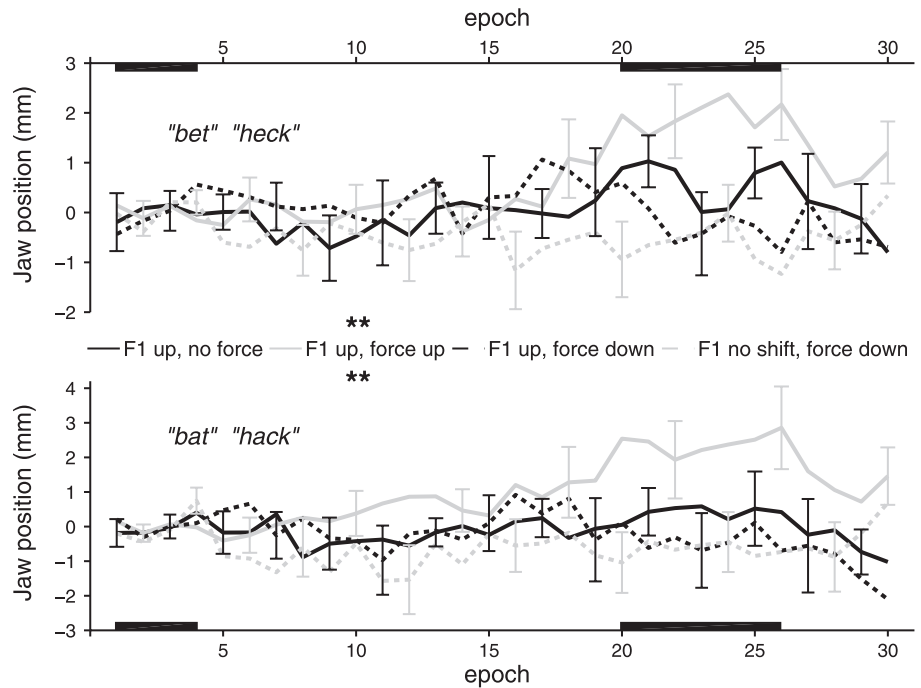


Fig. 5. Group mean jaw sensor position at peak displacement for /ε/ (words “bet” and “heck”) and /æ/ (words “bat” and “hack”) in four conditions. Individual subject data were normalized relative to the baseline mean value. Error bars indicate SEs. ***P* < 0.05 and Cohen’s *d* > 0.5 (see Table 5).

Generalization in Kinematic Dependent Variables

Generalization of after-effects in jaw and tongue position was examined for test words in the F1 no shift force down condition (the only condition without F1 shift). After a force perturbation, after-effects, and thus also the generalization of after-effects, would be reflected in a position deviation in the opposite direction of the perturbing force. However, as shown in Figs. 5 and 6, unlike the situation for the acoustic data, even the training words did not show any evidence of such after-effects in the kinematic measures. Consequently, position data for the jaw and tongue sensors during productions of the test words in the postperturbation phase also were not statistically different from those observed in the baseline phase.

Relation Between Auditory or Somatosensory Acuity and Sensorimotor Adaptation

Correlation coefficients for the relationship between either auditory accuracy (CE) or precision (DL) and the extent of adjustments in F1 during the F1 up no force condition were consistently very low and statistically nonsignificant (CE: *r* = -0.210 and *P* = 0.618 for /ε/ and *r* = -0.158 and *P* = 0.708

for /æ/; DL: *r* = -0.011 and *P* = 0.979 for /ε/ and *r* = 0.158 and *P* = 0.708 for /æ/). Similarly, only low and statistically nonsignificant correlations were found between either jaw somatosensory accuracy (CE) or precision (DL) and changes in jaw position during the F1 no shift force down condition (CE: *r* = 0.084 and *P* = 0.844 for /ε/ and *r* = -0.185 and *P* = 0.660 for /æ/; DL: *r* = -0.206 and *P* = 0.625 for /ε/ and *r* = -0.001 and *P* = 0.997 for /æ/). Consequently, the results from these correlational analyses will not be further discussed in the present report.

DISCUSSION

Adaptation to Only an Auditory Perturbation

When F1 in the auditory feedback was incrementally shifted up without simultaneous somatosensory perturbation, all subjects lowered F1 in their productions. This adaptive change decayed gradually rather than suddenly after the F1 perturbation was removed and partially generalized to untrained words (both untrained words with a vowel that differed from the training words and untrained words with consonants that differed from the training words showed a descriptive reduction in F1 during the postperturbation phase, although this change from baseline was not statistically significant for the latter type of words). These results are highly consistent with those from several previous formant shift studies (Houde and Jordan 1998, 2002; Max et al. 2003; Purcell and Munhall 2006a; Villacorta et al. 2007). One novel finding of the present study is the demonstration of selective adjustments in articulatory movements underlying such adaptive changes in F1. In particular, during word production in this F1 up no force condition, acoustic adaptation was associated with a raising of the tongue tip relative to the jaw so that the tongue tip reached a position that was higher than its typical position for these words. This adjustment reduced the amount of oral opening and, consequently, limited the perceived amount of upward shift of F1 in

Table 5. *t* and *P* values of paired *t*-tests and *d* values for normalized mandible height at movement offset for /ε/ and /æ/ in the baseline phase versus the max perturbation phase

	F1 Up, No Force	F1 Up, Force Up	F1 Up, Force Down	F1 No Shift, Force Down
/ε/				
<i>t</i> (df = 7)	1.287	2.647	-0.272	-1.138
<i>P</i>	0.239	0.033	0.793	0.293
<i>d</i>	0.5	0.9	0.1	0.4
/æ/				
<i>t</i> (df = 7)	0.555	2.402	-0.400	-1.010
<i>P</i>	0.596	0.047	0.701	0.346
<i>d</i>	0.2	0.9	0.1	0.4

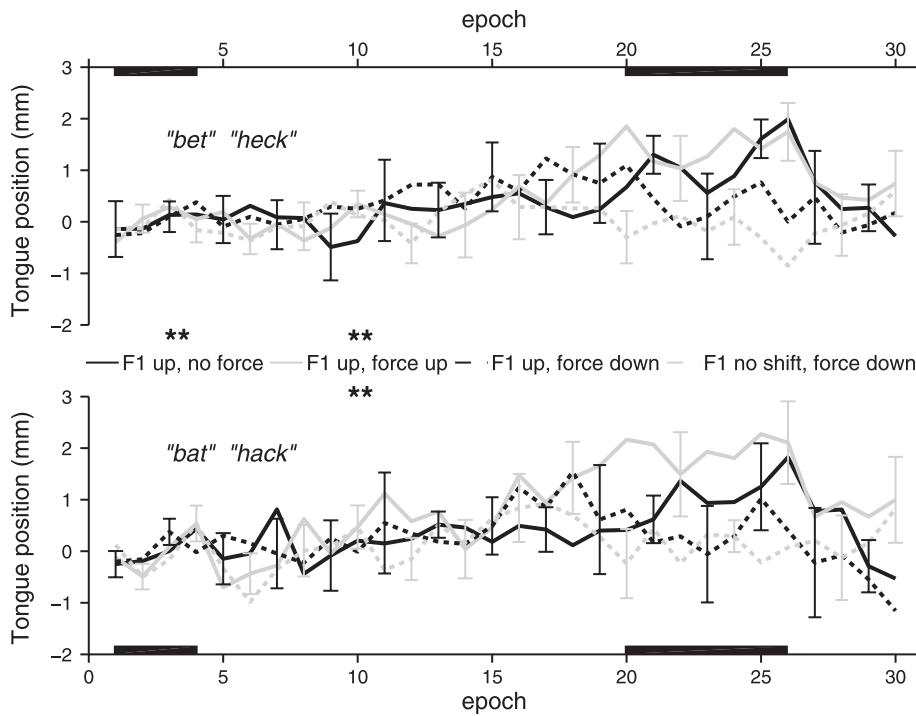


Fig. 6. Group mean anterior tongue sensor position at peak displacement for /ɛ/ (words “bet” and “heck”) and /æ/ (words “bat” and “hack”) in four conditions. Individual subject data were normalized relative to the baseline mean value. Error bars indicate SEs. ** $P < 0.05$ and Cohen’s $d > 0.5$ (see Table 6).

the auditory feedback given that oral opening and produced F1 are positively correlated in speech. Stated differently, decreasing the produced F1 by raising the tongue tip partially compensated for the experimentally increased F1 in the feedback signal. Hence, this kinematic adjustment reduced the mismatch between desired and perceived acoustic speech output, a first indication within these data to suggest that the auditory feedback stream plays a major role in the integration of multiple error signals provided by the different sensory modalities.

Further evidence supporting a major role for the auditory error signal is provided by two additional findings. First, in keeping with the results of Houde and Jordan (1998, 2002) and Villacorta et al. (2007), we found that adaptation in the acoustic aspects of speech generalized to some nonpracticed phonetic contexts. Here, generalization was explored by examining the generalization of after-effects. It cannot be determined from the present data set whether or not this insertion of test words among the postperturbation trials (when after-effects were already decaying) may account for the finding that generalization-related changes in F1 were statistically significant only for test words with a different vowel than the training

words and not for test words with the same vowel but different consonants. The influence, and possible interaction, of phonetic context and expired time since normal feedback was restored will need to be explored in further studies. Second, in the condition in which F1 in the auditory feedback was shifted up in the absence of a force perturbation, F2 showed a small but statistically significant increase. This finding replicates results reported by Villacorta et al. (2007) and is consistent with a previous proposal that speakers aim to achieve specific ratios of F1 and F2 in the acoustic domain (Miller 1989). In this F1 up no force condition, subjects lowered F1 in their productions, but because adaptation is incomplete, the net result is that the perceived F1 is still higher than during unperturbed productions. Consequently, F2 should also be raised, rather than lowered, to prevent too large a change in the ratio between F1 and F2.

Adaptation to Only a Somatosensory Perturbation

The experimental condition in which the jaw was perturbed by a force field but no simultaneous auditory perturbation was

Table 6. t and P values of paired t -tests and d values for normalized anterior tongue sensor height at movement offset for /ɛ/ and /æ/ in the baseline phase versus the max perturbation phase

	F1 Up, No Force	F1 Up, Force Up	F1 Up, Force Down	F1 No Shift, Force Down
<i>/ɛ/</i>				
t (df = 7)	2.684	2.721	0.661	-0.424
P	0.031	0.030	0.530	0.684
d	1.0	1.0	0.2	0.2
<i>/æ/</i>				
t (df = 7)	1.981	2.713	0.562	0.140
P	0.088	0.030	0.592	0.893
d	0.7	1.0	0.2	0.1

Table 7. t and P values of paired t -tests and d values for normalized decoupled anterior tongue sensor height at movement offset for /ɛ/ and /æ/ in the baseline phase versus the max perturbation phase

	F1 Up, No Force	F1 Up, Force Up	F1 Up, Force Down	F1 No Shift, Force Down
<i>/ɛ/</i>				
t (df = 7)	2.349	-0.602	1.234	0.739
P	0.051	0.566	0.257	0.484
d	0.8	0.2	0.4	0.3
<i>/æ/</i>				
t (df = 7)	2.198	-0.245	1.515	1.345
P	0.064	0.814	0.174	0.220
d	0.8	0.1	0.5	0.5

applied provides an opportunity to examine whether or not the jaw adapts to a downward mechanical force. Results from this F1 no shift force down condition indicate that, across trials, the jaw did indeed adapt to the incrementally increasing force so that it maintained an opening extent very similar to its typical extent for the target vowels. Similarly, the tongue also maintained its typical position. Thus, in the absence of a simultaneous auditory perturbation, neural control processes compensated for the mechanical perturbation and minimized or prevented error in the somatosensory modality. Based on this condition alone, however, it cannot be ruled out that the central nervous system keeps somatosensory error within limits merely because large kinematic deviations would result in failing to achieve desired acoustic consequences (note that an analogous interpretation does not apply to the condition with only an auditory perturbation because even not compensating at all would never lead to somatosensory error).

No kinematic after-effects were found when the force field was suddenly removed. This finding is in keeping with Nasir et al.'s (2006) finding that subjects compensated for a lateral position-dependent force during both consonant and vowel productions but without after-effects. Those authors hypothesized that force compensation was accomplished by impedance control (Nasir et al. 2006; see also Shiller et al. 2005). However, a prior study from the same laboratory made use of a velocity-dependent force in the direction of jaw protrusion and found very large after-effects (Tremblay et al. 2003). Consequently, the influence of perturbing force magnitude, direction (e.g., up-down vs. protrusion-retraction vs. lateral), and type (e.g., velocity dependent vs. displacement dependent) on the presence or absence of after-effects in jaw kinematics remains to be addressed in future studies.

Adaptation to Simultaneous Auditory and Somatosensory Perturbations

In two conditions, F1 was shifted up while simultaneously a downward or upward force was applied to the jaw. In one of these conditions (the F1 up force down condition), subjects' typical productions without compensation would result in compatible auditory and somatosensory error signals (indicating raised F1 and increased oral opening, respectively). For this condition, adapting to the somatosensory perturbation (resisting the force that acted to increase oral opening) was also beneficial for limiting the auditory error (which already indicated too much oral opening). In the other condition (the F1 up force up condition), subjects' typical productions without compensation would result in incompatible auditory and somatosensory error signals (indicating raised F1 and decreased oral opening, respectively). For this condition, not adapting to the somatosensory perturbation (not resisting the force that acted to decrease oral opening) was beneficial for adapting to the auditory error (which indicated too much oral opening).

The acoustic results show that subjects gradually decreased F1 in response to the upward shifted F1 in the auditory feedback in both the force up and force down conditions. In fact, in both cases, the time course and amount of acoustic F1 adaptation was descriptively similar to that for the F1 up no force condition. That is, F1 in the acoustic output exhibited consistent adaptation to F1-perturbed auditory feedback regardless of the presence or absence or the direction of a

simultaneous jaw perturbation. This finding contributes additional evidence in support of the notion that error monitoring in the auditory domain plays a more dominant role than error monitoring in the somatosensory domain.

In contrast to the F1 dependent variable, kinematic results based on jaw and tongue articulatory movements did not show the same result across the two conditions. The position of both the jaw and tongue at peak displacement increased in the F1 up force up condition, whereas the jaw maintained its normal position in the F1 up force down condition. Thus, subjects showed no adaptation, or at best minimal adaptation, to the force field and instead allowed the normal opening extent of the jaw and tongue to be reduced when auditory feedback suggested too much oral opening (F1 up) while somatosensory feedback suggested not enough jaw opening (force up). However, when both auditory and somatosensory feedback would have suggested too much oral opening if the force field was left uncompensated (F1 up force down), the jaw did show almost complete adaptation such that its normal opening extent was maintained. This finding that the jaw adapted to a force field when such adaptation facilitated adaptation to the acoustic perturbation but not when it would have opposed adaptation to the acoustic perturbation suggests once again that auditory error plays a more primary role than somatosensory error in the planning of speech articulation.

It is important to point out that the absence of a F1 no shift force up condition as a control condition (the current design of the study already required the subject's head to be immobilized for an hour with articulograph helmet, sensors, dental appliances, and insert earphones in place) does not affect this interpretation. If subjects were to not adapt to an upward force in the absence of an auditory perturbation, then, again, they did not take the somatosensory error into account during the planning of future movements. If they were to adapt to the force in those circumstances, then it still remains the case that the F1 up force up condition confirms that adjustments in planning based on somatosensory error were abolished when faced with incompatible auditory signals. Furthermore, it is also worth noting that the finding of reduced jaw opening in the F1 up force up condition cannot be accounted for by suggesting that subjects may have simply avoided the larger loads on the jaw (but note that in posttesting interviews all subjects indicated being unaware of the forces applied to the jaw) by actively decreasing the amount of oral opening. If this was true, they could have also used the same strategy in the F1 up force down condition.

Multimodality Sensory Integration in Speech Motor Control

Minimum-variance models. Some investigators have proposed linear weighted-sum systems to model the integration of multiple sensory information streams when estimating object properties or effector positions (Battaglia et al. 2003; Ernst and Banks 2002; Jacobs 1999; van Beers et al. 1999). This approach with additive models has been used, for instance, to model the integration of visual and proprioceptive signals in the estimation of hand position in a horizontal plane (van Beers et al. 1999). The integration of visual and proprioceptive signals in an additive manner requires the transformation of these signals to a common coordinate frame. In the common space, the visual modality senses the true state (X) as the visual estimate (X_v) with probability $P(X_v|X)$ and the proprioceptive

modality senses it as the proprioceptive estimate (X_p) with probability $P(X_p|X)$ (Ghahramani et al. 1997). The final estimate of state X is the summation of X_v with weight w_v and X_p with weight w_p , as follows: $\hat{X} = w_v X_v + w_p X_p$. Minimum variance estimation determines the weights by minimizing the variance of \hat{X} . Assuming that X_v is a Gaussian distribution with mean X and variance σ_v^2 and X_p is a Gaussian distribution with mean X and variance σ_p^2 , $w_v = \sigma_p^2 / (\sigma_v^2 + \sigma_p^2)$ and $w_p = \sigma_v^2 / (\sigma_v^2 + \sigma_p^2)$ (Ghahramani et al. 1997). Thus, in a given task, the more reliable sensory modality (smaller variance) is assigned a larger weight, whereas the less reliable sensory modality is assigned a smaller weight. It has been demonstrated that for the estimation of hand position vision dominates proprioception (Welch and Warren 1986). Although fixed relative weights have sometimes been assumed, other work has shown that the integration of sensory modalities may vary with context: estimation of hand position relied more on vision in the azimuth direction but more on proprioception in the depth direction (van Beers et al. 2002).

Linearly weighted integration of auditory and somatosensory information can account for only part of our results regardless of whether the common coordinate frame is in acoustic or spatial coordinates. The model can explain the results only if the controller has information about which modality caused the error signal in the common coordinates. In other words, unless the model can trace the auditory or somatosensory error and dynamically adjust the weight settings accordingly, some of the results from the present study cannot be accounted for. If auditory feedback was assigned a large weight and somatosensory feedback was assigned a small weight, F1 adaptation in the three F1 up conditions could be explained but not the almost-complete jaw adaptation in the F1 no shift force down condition. If auditory feedback was assigned a small weight and somatosensory feedback a large weight, the model could explain that the jaw maintained its typical opening in the two force down conditions and the no force condition but not that the jaw did not adapt to the upward force in the F1 up force up condition. If auditory and somatosensory feedback had equal weights, the model would predict different F1 adaptation extents in the three F1 up conditions, which is not consistent with our results. It has been previously reported that multisensory integration with fixed weights also fails to account for findings in the sensorimotor adaptation of arm movements with novel dynamics (Scheidt et al. 2005).

Extended minimum-variance models. More recently, it has been proposed that each sensory modality plays a dynamic role that is best represented by varying weights in an additive model. For example, Sober and Sabes (2003) manipulated visual feedback to investigate how vision and proprioception are integrated for planning reaching movements. The results suggested that visual feedback was used almost exclusively ($w_v = 0.97$) when deciding direction, whereas proprioceptive feedback dominated in determining motor commands ($w_p = 0.66$). In a follow-up study (Sober and Sabes 2005), subjects reached with visual feedback to either visual or proprioceptive targets. Simulations of the data indicated that the planning of movement direction relied strongly on vision for a visual target but approximately equally on vision and proprioception for a proprioceptive target. The subsequent phase of deriving motor commands for a movement in the chosen direction relied strongly on proprioception for both types of targets. Moreover,

when subjects reached to a proprioceptive target with visual feedback of the entire arm rather than only the fingertip, the role of vision in determining motor commands became equally large as that of proprioception.

Direct application of the model proposed by Sober and Sabes (2005) to the present findings is not feasible for two reasons. First, Sober and Sabes (2003, 2005) modeled the integration of sensory information during motor planning but without specifically considering this integration in the context of sensorimotor adaptation. One critical aspect of adaptation is that it depends on error signals from previous trials. Second, both visual and proprioceptive feedback were available during movement planning in the studies by Sober and Sabes (2003, 2005), but in our speech study there was no auditory feedback during movement planning due to the fact that each trial consisted of a monosyllabic word preceded by silence. Thus, only somatosensory information was available for estimation of the initial state of the vocal tract.

Nevertheless, following the suggestion by Sober and Sabes (2003, 2005) that weights for the involved sensory modalities can be flexibly adjusted, it can be hypothesized that, at least for the production of single words, 1) the planning of motor commands for speech depends entirely on somatosensory feedback to estimate the initial state, but 2) the movement targets are formulated in terms of a combination of articulatory positions that lie along an uncontrolled manifold (Schöner et al. 2008) resulting in equivalent acoustic/auditory characteristics that listeners perceive as the intended vowel. The feedforward control system that calculates motor commands may access an inverse internal model of the command-to-acoustics transformation that has been updated based on auditory feedback perceived during previous trials. In particular, this internal model is used to determine the uncontrolled manifold of articulatory positions that will accurately achieve the desired acoustic targets given the recently experienced sensorimotor mapping. Thus, the initial planning may be completed with a large weight assigned to information provided by the somatosensory system, but the inverse calculation of specific motor commands that will change the state from the initial condition to a desired target may be completed with a much larger weight assigned to information provided by the auditory system regarding the sensorimotor mapping experienced on previous trials. In this hypothesis, relevant auditory errors will lead to adaptation, whereas somatosensory errors will lead to adaptation only if the errors indicate that the estimated state does not lie along the uncontrolled manifold corresponding to the desired auditory consequences (as in the F1 up force down condition in the present study). If, on the other hand, somatosensory error is present together with auditory error, and the somatosensory error indicates a kinematic deviation toward a different uncontrolled manifold that would in fact yield an acoustic output that is now desirable given the simultaneously altered mapping between vocal tract configurations and acoustics, then no adaptation to the somatosensory error occurs (as in the F1 up force up condition).

Optimal feedback control. In this broader theoretical framework (Scott 2002, 2004; Todorov 2004; Todorov and Jordan 2002), an optimal feedback controller maps movement goals into motor commands by means of an internal inverse model, and a state estimator applies Bayesian optimal integration of actual sensory feedback, sensory consequences predicted by an

internal forward model, and prior knowledge. The optimal feedback controller uses the estimated state to update motor commands. Gains of the sensory systems are adjusted in a task-dependent manner such that errors are corrected only if they affect task performance, a principle known as minimum intervention (Todorov and Jordan 2002).

The question can be asked, then, which task performance index appears to have been used across the various conditions of the present study if the underlying controller is assumed to be an optimal feedback controller. In the F1 up no force condition, auditory feedback indicated that F1 was too high, whereas somatosensory feedback indicated that the articulatory positions were correct. In this condition, subjects elevated the tongue tip higher than its typical position and, thus, accepted somatosensory error to reduce auditory error, an observation consistent with a performance index defined in the auditory space. In the F1 no shift force down condition, the downward force on the jaw, if left uncompensated, would cause simultaneous auditory and somatosensory errors. The observation that the jaw compensated for the downward force and maintained its typical opening extent is compatible with either an auditory or somatosensory performance index. The same conclusion can be drawn for the F1 up force down condition in which the jaw again compensated for the downward force. In the F1 up force up condition, however, the force on the jaw remained uncompensated when the perturbation-induced change in jaw position was in the correct direction to compensate for the perceived auditory error. Consequently, the general pattern of our results is consistent with an optimal feedback controller that compensates for perturbations affecting a performance index defined in auditory space.

Conclusions and Future Directions

Combined, multiple converging findings from the present work suggest that, in speech production, sensorimotor adaptation updates underlying control mechanisms such that the future planning of vowel-related movements takes multiple error signals from previous trials into account but with a primary role for error in the auditory feedback modality. It remains unknown at this time how subjects would adapt to simultaneous perturbations of auditory and somatosensory feedback during consonant production. Because vocal tract constriction or obstruction is critical for consonant production, somatosensory information may be assigned a greater weight for consonant-related movements. In addition to exploring this question, future studies may quantitatively estimate the relative contribution of each modality. The weights of vision and proprioception for limb movement planning have been estimated by examining errors in movement direction and errors in generating motor commands (Sober and Sabes 2003, 2005). It may be possible to estimate the weights of auditory and somatosensory information for speech production by investigating acoustic or kinematic errors caused by perturbing the initial configuration of the vocal tract in combination with nonveridical auditory feedback from previous trials. The weights can then also be used to improve computational models of speech motor control. Finally, the experimental paradigm described here may yield interesting new insights into deficits in the neural mechanisms underlying disordered speech production, for example, by elucidating the foundation

for sensorimotor deficiencies in individuals who stutter (Max 2004).

ACKNOWLEDGMENTS

The authors thank Carol A. Fowler, Harvey R. Gilbert, and Mark Tiede for helpful suggestions.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC-007603 (Primary Investigator: L. Max PI).

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

DISCLOSURES

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

REFERENCES

- Abbs JH, Gracco VL.** Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. *J Neurophysiol* 51: 705–723, 1984.
- Battaglia PW, Jacobs RA, Aslin RN.** Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis* 20: 1391–1397, 2003.
- Baum SR, McFarland DH.** The development of speech adaptation to an artificial palate. *J Acoust Soc Am* 102: 2353–2359, 1997.
- Burnett TA, Freedland MB, Larson CR, Hain TC.** Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103: 3153–3161, 1998.
- Cohen HB.** Some critical factors in prism-adaptation. *Am J Psychol* 79: 285–290, 1966.
- Cohen J.** *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Hillsdale, NJ: Erlbaum, 1988.
- Ernst MO, Banks MS.** Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–433, 2002.
- Folkins JW, Abbs JH.** Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *J Speech Hear Res* 18: 207–219, 1975.
- Gescheider GA.** *Psychophysics: Method and Theory*. Hillsdale, NJ: Erlbaum, 1976.
- Ghahramani Z, Wolpert DM, Jordan MI.** Computational models of sensorimotor integration. In: *Self-Organization, Computational Maps and Motor Control*, edited by Morasso PG, Sanguineti V. Amsterdam: Elsevier, 1997, p. 117–147.
- Gracco VL, Abbs JH.** Variant and invariant characteristics of speech movements. *Exp Brain Res* 65: 156–166, 1986.
- Guenther FH, Ghosh SS, Tourville JA.** Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96: 280–301, 2006.
- Guenther FH, Espy-Wilson CY, Boyce SE, Matthies ML, Zandipour M, Perkell JS.** Articulatory tradeoffs reduce acoustic variability during american english /r/ production. *J Acoust Soc Am* 105: 2854–2865, 1999.
- Hamlet SL, Geoffrey VC, Bartlett DM.** Effect of a dental prosthesis on speaker-specific characteristics of voice. *J Speech Hear Res* 19: 639–650, 1976.
- Honda M, Fujino A, Kaburagi T.** Compensatory responses of articulators to unexpected perturbation of the palate shape. *J Phonetics* 30: 281–302, 2002.
- Houde JF, Jordan MI.** Sensorimotor adaptation of speech I: compensation and adaptation. *J Speech Lang Hear Res* 45: 295–310, 2002.
- Houde JF, Jordan MI.** Sensorimotor adaptation in speech production. *Science* 279: 1213–1216, 1998.
- Ito T, Kimura T, Gomi H.** The motor cortex is involved in reflexive compensatory adjustment of speech articulation. *Neuroreport* 16: 1791–1794.
- Jacobs RA.** Optimal integration of texture and motion cues to depth. *Vision Res* 39: 3621–3629, 1999.

- Jones JA, Munhall KG.** Learning to produce speech with an altered vocal tract: the role of auditory feedback. *J Acoust Soc Am* 113: 532–543, 2003.
- Max L, Wallace ME, Vincent I.** Sensorimotor adaptation to auditory perturbations during speech: acoustic and kinematic experiments. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by Solé MJ, Recasens D, Romero J. Barcelona: Universitat Autònoma de Barcelona, 2003, p. 1053–1056.
- Max L.** Stuttering and internal models for sensorimotor control: a theoretical perspective to generate testable hypotheses. In: *Speech Motor Control in Normal and Disordered Speech*, edited by Maassen B, Kent R, Peters HF, van Lieshout P, Hulstijn W. Oxford: Oxford Univ. Press, 2004, p. 357–388.
- Miller JD.** Auditory-perceptual interpretation of the vowel. *J Acoust Soc Am* 85: 2114–2134, 1989.
- Nasir SM, Ostry DJ.** Somatosensory precision in speech production. *Curr Biol* 16: 1918–1923, 2006.
- Perkell JS, Lane H, Ghosh S, Matthies ML, Tiede M, Guenther F, Ménard L.** Mechanisms of vowel production: auditory goals and speaker acuity. In: *Proceedings of the 8th International Seminar on Speech Production*, edited by Sock R, Fuchs S, Laprie Y. Strasbourg, France: 2008, p. 29–32.
- Perkell JS, Guenther FH, Lane H, Matthies ML, Perrier P, Vick J, Wilhelms-Tricarico R, Zandipour M.** A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *J Phonetics* 28: 233–272, 2000.
- Perkell JS, Matthies ML, Svirsky MA, Jordan MI.** Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: a pilot “motor equivalence” study. *J Acoust Soc Am* 93: 2948–2961, 1993.
- Purcell DW, Munhall KG.** Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am* 120: 966–977, 2006a.
- Purcell DW, Munhall KG.** Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119: 2288–2297, 2006b.
- Savariaux C, Perrier P, Orliaguet JP.** Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: a study of the control space in speech production. *J Acoust Soc Am* 98: 2428–2442, 1995.
- Scheidt RA, Conditt MA, Secco EL, Mussa-Ivaldi FA.** Interaction of visual and proprioceptive feedback during adaptation of human reaching movements. *J Neurophysiol* 93: 3200–3213, 2005.
- Schöner G, Martin V, Reimann H, Scholz J.** Motor equivalence and the uncontrolled manifold. In: *Proceedings of the 8th International Seminar on Speech Production*, edited by Sock R, Fuchs S, Laprie Y. Strasbourg, France: 2008, p. 23–28.
- Scott SH.** Optimal strategies for movement: success with variability. *Nat Neurosci* 5: 1110–1111, 2002.
- Scott SH.** Optimal feedback control and the neural basis of volitional motor control. *Nat Rev Neurosci* 5: 534–546, 2004.
- Shadmehr R, Mussa-Ivaldi FA.** Adaptive representation of dynamics during learning of a motor task. *J Neurosci* 14: 3208–3224, 1994.
- Shiller DM, Houle G, Ostry DJ.** Voluntary control of human jaw stiffness. *J Neurophysiol* 94: 2207–2217, 2005.
- Shiller DM, Sato M, Gracco VL, Baum SR.** Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am* 125: 1103–1113, 2009.
- Sober SJ, Sabes PN.** Flexible strategies for sensory integration during motor planning. *Nat Neurosci* 8: 490–497, 2005.
- Sober SJ, Sabes PN.** Multisensory integration during motor planning. *J Neurosci* 23: 6982–6992, 2003.
- Todorov E.** Optimality principles in sensorimotor control. *Nat Neurosci* 7: 907–915, 2004.
- Todorov E, Jordan MI.** Optimal feedback control as a theory of motor coordination. *Nat Neurosci* 5: 1226–1235, 2002.
- Tremblay S, Houle G, Ostry DJ.** Specificity of speech motor learning. *J Neurosci* 28: 2426–2434, 2008.
- Tremblay S, Shiller DM, Ostry DJ.** Somatosensory basis of speech production. *Nature* 423: 866–869, 2003.
- Vallabha GK, Tuller B.** Systematic errors in the formant analysis of steady-state vowels. *Speech Commun* 38: 141–160, 2002.
- van Beers RJ, Wolpert DM, Haggard P.** When feeling is more important than seeing in sensorimotor adaptation. *Curr Biol* 12: 834–837, 2002.
- van Beers RJ, Sittig AC, Gon JJ.** Integration of proprioceptive and visual position-information: an experimentally supported model. *J Neurophysiol* 81: 1355–1364, 1999.
- Villacorta VM, Perkell JS, Guenther FH.** Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am* 122: 2306–2319, 2007.
- Welch RB, Warren DH.** Intersensory interactions. In: *Handbook of Perception and Human Performance*, edited by Boff KR, Kaufman L, Thomas JP. New York: Wiley, 1986, p. 25.1.362306–25.
- Westbury JR, Lindstrom MJ, McClean MD.** Tongues and lips without jaws: a comparison of methods for decoupling speech movements. *J Speech Lang Hear Res* 45: 651–662, 2002.