Perceptual recalibration of speech sounds following speech motor learning

Douglas M. Shiller, Marc Sato, Vincent L. Gracco, and Shari R. Baum

Citation: The Journal of the Acoustical Society of America **125**, 1103 (2009); doi: 10.1121/1.3058638
View online: https://doi.org/10.1121/1.3058638
View Table of Contents: https://asa.scitation.org/toc/jas/125/2
Published by the Acoustical Society of America

Perceptual recalibration of speech sounds following speech motor learning

Douglas M. Shiller[a)]
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada and Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada

Marc Sato
Départment Parole et Cognition, GIPSA-lab, UMR CNRS 5216, Grenoble Universités, 1180 Avenue Centrale, BP 25, 38040 Grenoble Cedex 9, France

Vincent L. Gracco
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada; Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada; and Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut 06511

Shari R. Baum
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada and Centre for Research on Language, Mind and Brain, McGill University, 3640 de la Montagne, Montreal, Quebec H3G 2A8, Canada

(Received 6 June 2008; revised 26 November 2008; accepted 8 December 2008)

The functional sensorimotor nature of speech production has been demonstrated in studies examining speech adaptation to auditory and/or somatosensory feedback manipulations. These studies have focused primarily on flexible motor processes to explain their findings, without considering modifications to sensory representations resulting from the adaptation process. The present study explores whether the perceptual representation of the /s-ʃ/ contrast may be adjusted following the alteration of auditory feedback during the production of /s/-initial words. Consistent with prior studies of speech adaptation, talkers exposed to the feedback manipulation were found to adapt their motor plans for /s/-production in order to compensate for the effects of the sensory perturbation. In addition, a shift in the /s-ʃ/ category boundary was observed that reduced the functional impact of the auditory feedback manipulation by increasing the perceptual "distance" between the category boundary and subjects' altered /s/-stimuli—a pattern of perceptual adaptation that was not observed in two separate control groups. These results suggest that speech adaptation to altered auditory feedback is not limited to the motor domain, but rather involves changes in both motor output and auditory representations of speech sounds that together act to reduce the impact of the perturbation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3058638]

PACS number(s): 43.70.Mn, 43.71.Es, 43.70.Bk [AL] Pages: 1103–1113

I. INTRODUCTION

The functional sensorimotor nature of speech production has been demonstrated in a number of studies employing manipulations of both somatosensory and auditory feedback. Introducing unexpected auditory or somatosensory perturbations during speech production results in rapid (online) *compensatory* motor changes (Abbs and Gracco, 1984; Gracco and Abbs, 1985; Kawahara, 1995) while using more predictable and constant changes in auditory or somatosensory feedback results in a *recalibration* (or relearning) of the mapping between sensory signals and motor output (Baum and McFarland, 1997; Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2003; Tremblay *et al.*, 2003; Nasir and Ostry, 2006; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*, 2007). For example, in a seminal investigation by Elman (1981), participants were presented with normal or frequency-shifted auditory feedback in real time during a task requiring them to instantaneously reproduce (i.e., shadow) shifts in fundamental frequency (F0) in synthetic vowel stimuli presented to them. Results revealed that in the frequency-shifted condition, participants shifted their F0 production in a compensatory fashion in an effort to achieve a target F0 in the feedback they received. A comparable compensatory effect was demonstrated using sentential stimuli (Elman, 1981).

Since that early study, numerous investigations have made use of real-time alterations in auditory feedback to explore resulting output characteristics (as well as patterns of generalization), including studies of shifts in fundamental frequency (F0—Kawahara, 1995; Jones and Munhall, 2000, 2005) and formant frequencies (Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*,

[a)]Author to whom correspondence should be addressed. Electronic mail: doug.shiller@mail.mcgill.ca

2007) during vowel production. All of these investigations have reported significant (but incomplete) compensatory effects in the speech motor output, as measured via kinematic or acoustic means, suggesting—not surprisingly—a strong link between auditory perception and motor output. Importantly, these studies have demonstrated not only motor corrections to counteract the effect of the perturbation, but also a persistence of those corrections (i.e., an aftereffect) once the perceptual manipulation is removed (e.g., Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a, 2006b; Villacorta *et al.*, 2007). The fact that changes in the motor system do not disappear immediately likely reflects a change in motor representations due to a global remapping of the auditory-motor relationship (i.e., sensorimotor adaptation). These results are typically interpreted in the framework of forward and inverse internal models (Guenther, 1995; Perkell *et al.*, 1997; Kawato, 1999). Modulated response outputs during the manipulation of somatosensory or auditory feedback are thought to reflect feedback control mechanisms in which the expected sensory consequences of the speech-motor act are evaluated against the actual sensory input in order to further control production. These mechanisms also help to distinguish the sensory consequences of our own actions from sensory signals due to changes in the outside world (see Guenther, 2006, for a review).

Although these findings suggest an important relationship between an individual's perceptual "space" and his/her speech motor patterns, they fail to reveal the extent to which the perception and production systems may be truly integrated in producing such adaptation effects. One recent study attempted to draw a stronger association between perception and production, by examining how individual differences in auditory discrimination abilities may influence the degree to which speakers adapt to altered sensory feedback (Villacorta *et al.*, 2007). The investigation demonstrated that those individuals who exhibited greater sensitivity in auditory discrimination of relevant acoustic features (i.e., F1 frequencies) produced greater degrees of compensation to alterations in F1 in real-time auditory feedback. However, an important issue that has not been addressed is whether speech-motor adaptation to alterations in sensory feedback modifies the perceptual representation of speech sounds. Indeed, while the above-mentioned studies of speech adaptation to altered sensory feedback have indicated that sensory input affects speech motor control, the extent to which speech motor processes influence the speech sound representations that are central to how we perceive and produce speech has not been previously explored. That is, with changes in speech-motor output to alterations in auditory feedback, is there a concomitant adjustment in the perceptual space?

A significant body of speech perception research has demonstrated that sensory representations of speech sounds are flexible in response to changes in the sensory and linguistic aspects of speech input (Ladefoged and Broadbent, 1957; Miller and Liberman, 1979; Mann and Repp, 1980; Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005). Listeners have been shown to rapidly compensate for short-term changes in speaking rate (Miller and Liberman, 1979) and phonetic context (Mann and Repp, 1980, 1981), as well as individually varying vocal tract characteristics (i.e., speaker normalization (Ladefoged and Broadbent, 1957; Nearey, 1989), in order to maintain perceptual accuracy in the face of such variability. Perceptual learning has been observed following exposure to foreign-accented talkers, resulting in improved word identification performance over time (Clarke and Garrett, 2004; Bradlow and Bent, 2008). In addition, a number of studies have shown changes in perceptual speech sound representations in listening tasks involving lexically ambiguous stimuli (Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006).

In spite of this significant body of speech perception research demonstrating that sensory representations of speech sounds are flexible in response to changes in the sensory and linguistic aspects of speech input, studies of adaptation in speech production have focused primarily on the flexibility of motor processes in order to explain their findings, without regard for the possible contribution of changes in sensory representations that are presumed to constitute the acoustic "target" of speech movements. In the present study, we utilized a speech adaptation paradigm to explore whether auditory representations of speech sounds are, in fact, not static following speech production under conditions of altered auditory feedback, but rather can be adjusted to reflect changes in a talker's own speech output. The experimental procedure involved the real-time alteration of auditory feedback during the production of the sibilant /s/ in brief /s/-initial words. During an intensive period of speech practice under feedback-altered conditions, /s/-productions were examined for evidence of compensation for the manipulation. The persistence of any change in output following the sudden removal of the perturbation was also explored, as a reflection of a change in the speech-motor representation (Baum and McFarland, 1997; Houde and Jordan, 1998; Baum and McFarland, 2000; Houde and Jordan, 2002; Jones and Munhall, 2003; Tremblay *et al.*, 2003; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In addition to motor adaptation, the perceptual representation of the /s-ʃ/ contrast was examined immediately prior to and following the period of speech adaptation using a phoneme labeling task in order to determine whether boundary shifts would emerge, suggesting a perceptual adaptation to reduce the functional impact of the auditory feedback alteration.

II. METHODS

All procedures were approved by the Institutional Review Board of the Faculty of Medicine at McGill University and all subjects provided informed consent prior to testing.

A. Subjects

Thirty subjects were tested, all females (in order to reduce between-subject variability in fricative centroid frequency and vowel F0), between 19 and 30 years of age, and native speakers of North American English. Subjects had no reported history of speech or language disorder and no hear-

ing loss (confirmed by a pure-tone hearing screening conducted prior to testing). Subjects were also evaluated informally to rule out the presence of a functional speech disorder involving the production of /s/ (i.e., interdental, dentalized, lateral, or palatal lisp) by one of the authors (Shiller), who is a certified speech-language pathologist. The 30 subjects were randomly assigned to one of three groups, with ten subjects in each group: (1) a group that produced speech under conditions of *altered* auditory feedback (group AF), (2) a group that produced speech under conditions of *unaltered* auditory feedback (group UF), and (3) a group that passively listened to a sequence of frequency-altered speech stimuli that was matched to the stimuli perceived by subjects in the AF group, only in this case without speaking (PL group).

B. Audio recording

All groups were tested while seated in a sound attenuating testing room (Industrial Acoustics Company, Bronx, NY). For tasks involving speech production (AF and UF groups), subjects spoke into a condenser microphone (ME-66, Sennheiser, Germany) positioned 10 cm from the subject's mouth. The signal from the microphone was amplified to line level using a microphone preamplifier (model 302, Symetrix, Mountlake Terrace, WA) and then passively split into two identical channels, one of which was used to digitally record the subject's unprocessed speech signal, and the other which was sent to a digital signal processor (DSP) for processing. The output of the DSP was again passively split into two identical channels, one of which was digitally recorded (simultaneously with the unaltered speech signal) and the other which was presented back to the subject through circumaural headphones (SR-80, Grado Labs, Brooklyn, NY). The two channels of audio (unprocessed and processed by the DSP) were analog low-pass filtered at 22.0 kHz, and then digitized at 44.1 kHz (16 bit resolution) using an analog-to-digital converter (Transit, M-Audio, Irwindale, CA) attached via USB to a Toshiba laptop computer. The digitized audio signals were captured directly into MATLAB (v.7.4, Mathworks, Natick, MA) using the DATA ACQUISITION TOOLBOX (v. 2.10, Mathworks, Natick, MA).

C. Manipulation of auditory feedback

Subjects in the AF group produced a sequence of individual words under conditions of altered auditory feedback. Auditory feedback was manipulated using a DSP to gradually modify in real time the frequency spectrum of short words containing the initial fricative /s/ (e.g., "see"). Under conditions of maximal acoustic perturbation, the DSP shifted the first spectral moment (or centroid) of the fricative down by three semitones (averaging −1430 Hz across subjects), resulting in an acoustic signal that, while still categorically an /s/, was now closer in centroid frequency to the fricative /ʃ/ (as in "she"). The vowel spectrum was shifted to the same degree (reducing the fundamental frequency and all formants), which had the effect of lowering the perceived pitch of the voice. This modified acoustic signal was amplified and fed back to subjects through headphones.

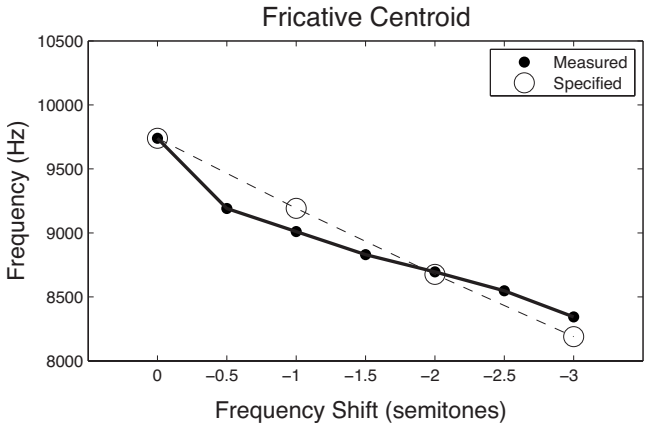

FIG. 1. Empirical test of the ability of the DSP to alter fricative centroid frequency. The dashed line (open circles) indicates the specified magnitude of spectral shifting (in steps of −0.5 semitones). The solid line (filled circles) indicates the measured centroid frequency following processing by the DSP. This test confirms the ability of the DSP to significantly reduce the centroid frequency of *s* within a reasonably small margin of error (<5% at the maximum specified shift of −3.0 semitones).

The DSP used to perform the real-time manipulation of auditory feedback was a commercial device designed for the manipulation of speech acoustic signals (SPX-1000, Yamaha, Japan). The device digitally samples an input signal at 44.1 kHz and uses a proprietary method to "pitch shift" an input speech signal at a delay of 10 ms. The ability of the device to manipulate the parameter of interest in the present study—fricative first spectral moment—was assessed empirically in order to verify the operation of the device. A speech utterance "see" was recorded on a digital audio tape (44.1 kHz, 16 bit sampling) and then played back through the DSP with the device set at 0 frequency shift (baseline, unmodified signal) and then at six linearly decreasing steps of frequency shift, ranging from −0.5 to −3.0 semitones (the maximum employed in the present study) in steps of −0.5 semitones. For each level of spectral shift, the output of the DSP was digitally recorded and the fricative centroid frequency was estimated using the same method as was employed for the analysis of speech production in the study (see Sec. II H). Figure 1 shows the recovered values of fricative centroid frequency at each level of frequency shift (the open circles show the specified shift and the filled circles indicate the recovered frequency). A close correspondence was observed between the specified and recovered frequencies; at the maximum level of spectral shifting (−3.0 semitones, which corresponds to the manipulation employed in the present study), the deviation between specified and recovered centroids was less than 5%. This was satisfactory for the purpose of the present study, as it confirmed the ability to significantly reduce the fricative centroid frequency with a reasonably small margin of error in the magnitude of the specified shift.

In previous studies of speech adaptation involving manipulations of auditory feedback during vowel production, masking noise has been mixed with the subject's altered speech signal before presenting it back to them through headphones in order to minimize the subject's perception of their unmodified speech output via air and bone conduction.

TABLE I. Speech production stimuli.

/s/-stimuli	sue, see, saw, sigh, say, sot, so, seep, sip, sop, sock, suit, seat, soy, soup
/ʃ/-stimuli	shoe, she, shaw, shy, shay, shot, show, sheep, ship, shop, shoot, sheet, shape, shake, shut

In the present study, no masking noise was added to the modified speech signal due to its potential impact on the noise spectrum of the voiceless fricatives that were of primary interest. Rather, the modified acoustic signal was simply gained sufficiently to limit the subjects' perception of their air/bone conducted speech signal. The amplifier gain levels and speaking volume employed in the present study were determined in pilot tests during which subjects were instructed to produce words at a comfortable speaking volume while the gain levels on the microphone preamplifier and DSP were adjusted. The goal was to achieve a sound level that was perceived by the subject to be loud (but not uncomfortable), that was reported by subjects to limit the perception of their own air/bone conduction speech signal, and that was found to yield clear evidence of speech adaptation in the fricative. With the microphone positioned at a fixed distance of 10 cm from the subjects, a gain level that achieved these goals was determined, at which point the peak level indicated on the volume unit (VU) meter of the microphone amplifier was noted (+2, on a scale from −7 to +3). For all subsequent subjects, with the variable gain settings on all sound equipment fixed, a consistent loudness level was achieved by ensuring that the subjects' mouths were positioned 10 cm from the microphone and that their speech yielded a target peak VU level of +2 [approximately 65 dB sound pressure level (SPL), as measured at the microphone using a handheld SPL meter]. Feedback regarding speaking volume was provided to subjects during a brief practice period prior to testing, as well as throughout the course of testing (using visual feedback from the experimenter), as required.

During the experiment, the DSP was controlled by the laptop computer using an external USB MIDI interface (MIDIsport 2×2, M-Audio). Control of the DSP was coordinated with the presentation of visual stimuli and audio recording using custom software written in MATLAB.

D. Task sequence: AF and UF groups (speech production)

For the AF and UF groups, the speech production task involved reading a sequence of words, presented one at a time on a computer display (65 point font) at a distance of 1.5 m. Each word was presented on the display for 3 s, followed by a 1 s period in which the display was blank, for a total interstimulus interval (onset to onset) of 4 s. Stimuli consisted of single-syllable English words of the form consonant-vowel (CV) or consonant-vowel-consonant (CVC). The total set of stimuli consisted of 30 words (Table I) 15 of which had /ʃ/ as the onset sound (not used for speech training, as described below) and 15 of which had /s/ as the onset sound (used for speech training). The stimulus word set included a range of vowel sounds; however, the final consonant sounds were restricted to the unvoiced stops: /p/, /t,/ or /k/ (see Table I). Subjects were instructed to read the words with the initial sound prolonged, following a model provided by the experimenter. In pilot studies, prolonged fricatives tended to show less overall variability in duration and amplitude as well as a more consistently present steady-state region at the fricative center. The modification also served to increase the subjects' exposure to the auditory feedback manipulation (in the AF group). The mean duration of baseline /s/-productions was 599 ms.[1]

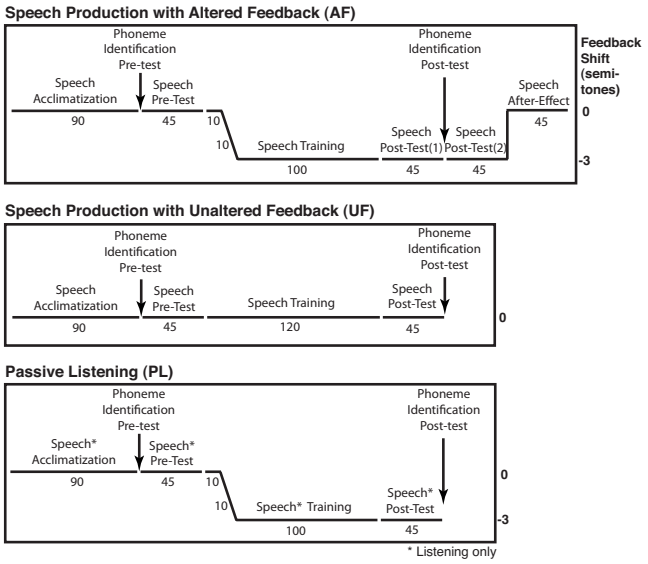

FIG. 2. Schematic depicting the sequence of procedures for each of the three groups of subjects: (1) speech production with altered auditory feedback (AF, top), (2) speech production with unaltered auditory feedback (UF, middle), and (3) passive listening (PL, bottom). The numbers underneath the horizontal lines indicate the number of words spoken. See text (Secs. II D and II E) for details.

For both the AF and UF groups, subjects underwent the following sequence of six procedures (see Fig. 2 for schematic).

(1) *Acclimatization period*. Subjects read aloud 90 words into a microphone while listening to their amplified (but otherwise unaltered) speech acoustic signal through headphones. The stimuli consisted of an equal proportion of words beginning with /s/ and /ʃ/, drawn from the full set of 30 stimulus items. Each word was presented three times in a fully randomized order.
(2) *Phoneme identification pretest*. Following the acclimatization period, subjects underwent the first of two phoneme identification tasks, which involved listening to synthetic speech stimuli through headphones and assigning a phoneme label to each token by responding on a computer keyboard (see *Phoneme identification* below).
(3) *Speech production pretest*. Subjects then underwent an assessment of speech production (the first of four for the AF group and the first of two for the UF group). The assessment involved the production of a restricted set of CV speech stimuli, consisting of the fricative /s/ combined with the three English "point" vowels: /u/ (as in "sue"), /i/ (as in "see") and /a/ (as in "saw"). Each word was produced ten times in a fully randomized order.

While the fricative /s/, produced in a range of vowel contexts, was the primary focus of the assessment, an additional 15 tokens involving the fricative /ʃ/ were also included in the assessment in order to evaluate the possible generalization of *s*-production training to /ʃ/. Due to space limitations, the analysis of /ʃ/-words is not reported in the present paper.

(4) *Speech training*. Subjects produced a sequence of words containing the fricative *s* exclusively (drawn from the full set of 15 /s/-stimuli). For the AF group, /s/-words were produced under conditions of altered auditory feedback, whereas for the UF group, /s/-words were produced with similarly amplified but otherwise unaltered auditory feedback. For the AF group, this training period began with 10 trials under unaltered feedback conditions, followed by the introduction of the acoustic perturbation (linearly ramped on over 10 trials), and then 100 trials under conditions of maximal acoustic perturbation (−3.0 semitones). For the UF group, the training period consisted of 120 trials under conditions of unaltered auditory feedback.

(5) *Speech production post-test (1)*. Both groups then underwent a second assessment of speech production consisting of the same 45 stimuli that were used in the speech production pretest (with the *s*-stimuli presented in a different random order). For both groups, auditory feedback conditions for this assessment remained unchanged from the preceding training period. Hence, for the AF group, subjects continued to experience the maximum level of acoustic perturbation, while for the UF group, feedback remained unaltered. For both groups, motor adaptation was assessed as the difference in /s/-centroid frequency between this test and the speech production pretest (item 3 above).

(6) *Phoneme identification post-test*. Following the speech production post-test, subjects in both groups underwent a second phoneme identification procedure (same as procedure 2 described above), but utilizing a different randomized order of perceptual stimuli.

For subjects in the AF group, two additional procedures were carried out following the phoneme identification post-test in order to examine the presence of a motor learning aftereffect.

(7) *Speech production post-test (2)*. A replication of the speech production post-test (item 5 above) was carried out (using a new randomized stimulus order) under conditions of maximal auditory feedback perturbation in order to ensure that any changes in speech output following speech production training were maintained during the phoneme identification post-test (a 7–8 min period during which the subject listened to speech stimuli without speaking).

(8) *Speech production aftereffect*. Immediately following the replicated speech production post-test (2), the perturbation of auditory feedback was suddenly and unexpectedly removed and a final assessment of speech production was carried out under conditions of unaltered feedback. Once again, the assessment consisted of the same stimuli as used in all previous assessments of speech production, but in a different randomized order.

E. Task sequence: PL group (passive listening)

Subjects in the PL group were seated in the sound-attenuating testing room and viewed the presentation of individual words on a computer display (using same text size, timing, and distance as described above). Simultaneous with each visual stimulus, a corresponding spoken word was presented auditorily through headphones.

The auditory stimuli were tokens digitally recorded from the output of the DSP for one subject who had participated in the study as a member of the AF group. The subject was selected on the basis of the following criteria: (1) a degree of /s/ motor adaptation that was similar to, but did not exceed, the average degree of adaptation for the entire AF group, and (2) a degree of token-to-token variability that did not exceed the average for the AF group. The subject that was selected exhibited a mean /s/-adaptation effect of 358 Hz (group mean: 527 Hz) and an across-token standard deviation of /s/-centroid frequency that ranged from 354.4 Hz in the baseline phase to 300.9 Hz at the end of training (group means: 503.9 and 452.4 Hz, respectively).

Following the same sequence as the AF group, subjects in the PL group first listened to 90 "acclimatization" trials and then underwent the phoneme identification pretest, using the same procedure administered to the AF and UF groups. Subjects then listened to the speech production pretest (45 trials), speech training (10 unaltered, 10 ramping on, and then 100 trials with maximal spectral shifting), and speech production post-test (45 trials at maximal spectral shift), after which they underwent a second test of phoneme identification (phoneme identification post-test).

In order to ensure that subjects in the PL group attended to the auditory and visual stimuli, they participated in a simple task in which they were instructed to indicate (using a computer keyboard) the number of letters contained in each word that was presented on the screen. Because the auditory presentation typically followed the visual presentation by approximately 1 s, subjects were instructed to respond with a key press only after the word had been presented auditorily.

F. Phoneme identification

A phoneme identification procedure was carried out in order to characterize the boundary between each subject's /s/ and /ʃ/ categories. The task employed a set of synthetic speech stimuli, which differed from each other along a nine-step acoustic spectral continuum from /s/ to /ʃ/. In the task, individual speech utterances, consisting of the sound embedded within the carrier: "a _ed" (e.g., "a shed" or "a said") were presented to subjects at a comfortable volume through headphones. Following each stimulus presentation, subjects labeled the fricative by pressing a corresponding key (labeled "s" or "sh") on a computer keyboard using the index and middle fingers of their dominant hand. Subjects were instructed to respond as quickly and accurately as possible following the onset of the stimulus. Key order was counterbalanced such that in each group, half of the subjects used each order. In total, ten tokens of each stimulus were pre-

sented to each subject in random order. Ten practice trials (randomly selected) were added at the beginning of each labeling session, yielding a total of 100 tokens per test (10 practice+90 test).

G. Synthetic speech stimuli

The synthetic speech stimuli used in the phoneme identification task were the same as those used in a recent study by Lane *et al.* (2007); thus, a detailed description and further references may be found in that paper. Briefly, speech stimuli were synthesized on the basis of natural exemplars of the phrases: "a said" and "a shed," produced by a female speaker. Two synthetic fricative segments were produced using a Klatt formant synthesizer, such that the frequency and amplitude of the formants (spectral peaks) were closely matched to the formants of the naturally produced fricatives. The parameters used to generate the two synthetic segments (/s/ and /ʃ/) were then adjusted slightly so that they differed only in terms of formant amplitude (i.e., formant frequencies were aligned). Finally, using these two synthetic fricatives as the boundary stimuli, a /s-ʃ/ continuum was generated by interpolating the formant amplitudes over seven intermediate steps. The interpolation was not perfectly linear, due to the constraint of having to use integer values to specify formant amplitudes in the Klatt synthesizer, as well as the desire for a continuum that was roughly balanced for the number of stimulus steps on either side of the category boundary. Finally, the remaining portions of the utterance (preceding vowel "a" and following coda "-ed") were concatenated with the fricative to yield the complete stimuli (e.g., "a said") (see Lane *et al.*, 2007 for further details).

H. Data analysis

1. Acoustics

Adaptation of *s*-production was assessed by examining changes in the first spectral moment: a stable acoustic property of fricatives (Behrens and Blumstein, 1988; Jongman *et al.*, 2000) that has been used to evaluate the accuracy of *s*-production in a number of studies involving the manipulation of sensory feedback (Baum and McFarland, 1997, 2000). The first spectral moment, or centroid frequency, is a measure of central tendency in the spectral domain and is computed as the amplitude-weighted mean of the frequency spectrum (obtained by discrete Fourier transform). For each utterance, the mean frequency centroid was obtained for a 100 ms window about the midpoint of the fricative.

In order to evaluate the magnitude of changes in fricative production within each group of subjects, mean centroid estimates were first obtained separately for each subject. For each of the speech assessments (four for the AF group, and two for the UF group—as described above), an average centroid was computed across all /s/-utterances (collapsing across the three vowel contexts[2]). Following the calculation of these within-subject mean values, difference scores were computed between relevant assessments. Two difference scores, in particular, were of interest: First, for all subjects in the AF and UF groups, the difference between the pretraining (baseline) and post-training (immediately following the training period) assessments was obtained in order to measure the direction and magnitude of the speech training effect; second, for all subjects in the AF group, the difference between the initial baseline assessment and the final assessment (immediately following removal of the shift) was obtained in order to measure the direction and magnitude of the speech adaptation aftereffect (the persistence of motor learning effects in the absence of auditory perturbation). Statistical pairwise comparisons were carried out using *t*-tests, corrected for multiple comparisons (familywise $p<0.05$) using Holm's sequential Bonferroni procedure.

2. Phoneme identification function

The set of response data from the phoneme identification task (selection of "s" or "sh" for ten tokens of nine different stimuli) was used to estimate parameters for the /s-ʃ/ identification function for the two tests (pre- and postpractice) within each subject. This was done by first computing the proportion of "s" responses for each stimulus (1.0=100% "s" response), linearly interpolating to an interval of 0.1 stimulus steps, and then fitting a four-parameter logistic function (sigmoid) to the resulting data points. The sigmoid parameters include the stimulus step at which the proportion of "s" responses is 0.5, which was taken as the boundary between the "s" and "sh" responses. As in the case of the centroid frequency measure, estimates of the location of the sigmoid boundary from the two assessments (pre- and post-training) were converted to a difference score for each subject in order to measure the direction and magnitude of the training effect. Pairwise comparisons between the AF and other groups were carried out using independent-samples t-tests, corrected for multiple comparisons (familywise $p<0.05$) using Holm's sequential Bonferroni procedure.

III. RESULTS

On average, the auditory feedback manipulation resulted in a 1430 Hz reduction in /s/-centroid frequency across subjects in the AF group. Following the period of speech practice under these altered auditory conditions, these subjects adjusted their production of /s/ to counteract the effect of the sensory perturbation (Fig. 3, left bar). Specifically, subjects showed an increase in /s/-centroid frequency (i.e., in the direction opposite that of the spectral shift) averaging 529 Hz following training. The mean change in fricative centroid frequency within each of the three tested vowel contexts (/u/, /i/, and /a/) was found to vary somewhat (averaging 567.5, 462.7, and 556.0 Hz, respectively); however, the difference between contexts was not statistically reliable ($F[2,18]=0.542$, $p>0.05$). The overall change in speech output was found to persist following the sudden removal of the perturbation (Fig. 3, middle bar), averaging 490 Hz across a final block of 30 /s/-word trials with unaltered auditory feedback. Note that a deadaptation effect of 113 Hz was observed between the first five and final five /s/-words in this final testing block; however, the reduction was not statistically reliable [$t(9)=0.49$, $p>0.05$]. The persistence of motor adaptation following removal of the sensory perturbation indicates that the changes in speech production were not the result of on-

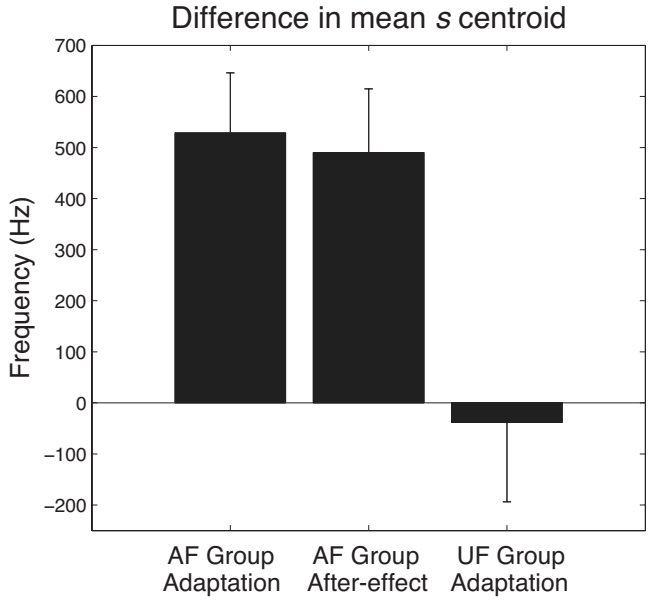

FIG. 3. Observed changes in speech output. The observed changes in /s/-*centroid* frequency associated with speech practice under conditions of altered auditory feedback (AF group; *left* and *middle* bars) and under conditions of unaltered auditory feedback (UF group; *right bar*). The error bars show one standard error. See text for details.

line corrections, but rather resulted from changes in the neurally specified motor plan for the production of the /s/-sound. The observed adaptation and aftereffects for the AF group were found to be reliably larger than the change in /s/-production observed in a group of control subjects (UF group) who underwent an identical period of intensive speech practice with *s*-initial words under conditions of *unaltered* auditory feedback [AF adaptation versus UF adaptation: $t(18)=2.9$, $p<0.05$; AF aftereffect versus UF adaptation: $t(18)=2.7$, $p<0.05$]. For this control group, little change in speech output was observed following speech practice [mean: −39 Hz, $t(9)=0.25$, $p>0.05$], indicating that the effects observed in the experimental group did not arise from aspects of the experimental setup unrelated to the spectral manipulation (such as the high concentration of /s/-words in the speech corpus) (Fig. 3, right bar).

In addition to assessing changes in /s/-production resulting from the auditory feedback manipulation, an evaluation of speech perception was carried out for both the AF and UF groups prior to and following the period of intensive /s/-practice in order to examine changes in the representation of speech-sound categories resulting from the manipulation. Mean /s-ʃ/ identification functions and their associated boundary locations (the stimulus step at which the proportion of "s" responses is 0.5) are shown in Figs. 4(a) and 4(b). Following the period of /s/-word practice with altered auditory feedback (in which the fricative centroid frequency was shifted lower, toward /ʃ/), subjects in the AF group exhibited a change in their /s-ʃ/ identification function that acted to reduce the impact of the perturbation. Specifically, the location of the /s-ʃ/ category boundary was shifted toward a lower centroid frequency (in the direction of the /ʃ/ category),

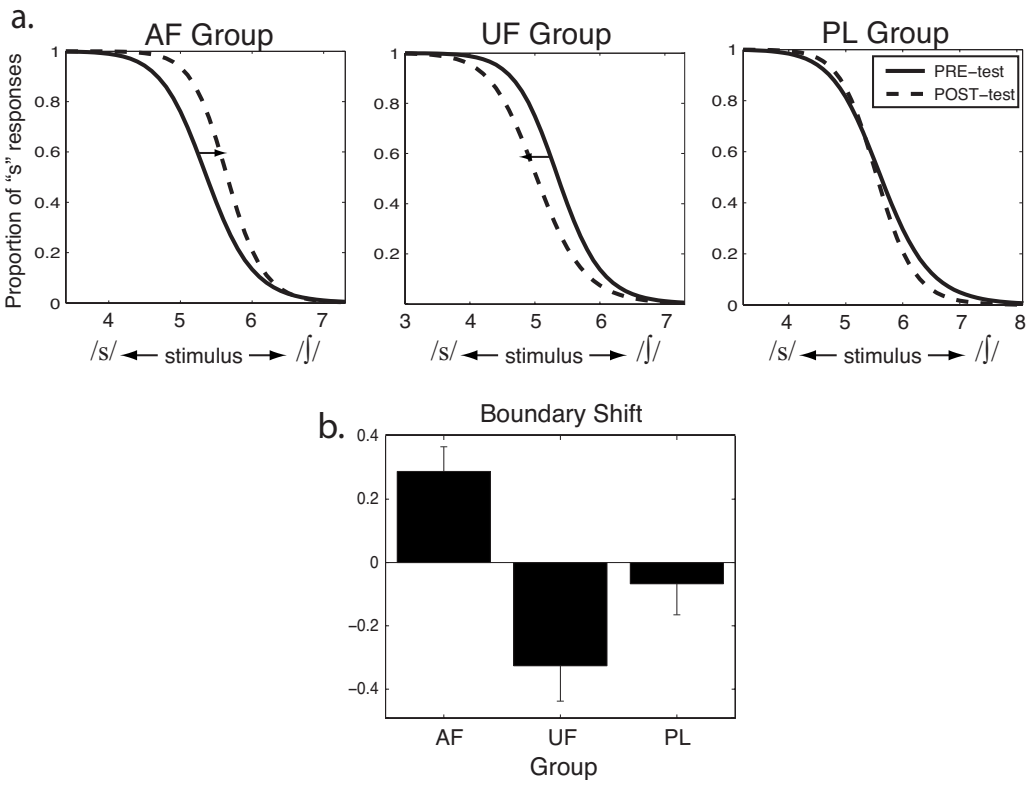

FIG. 4. Evaluation of speech-sound perception. (a) Mean pretest and post-test identification functions for each group, based on average sigmoid boundary and slope parameters (see Sec. II H). Notably, for subjects in the AF group, the boundary is seen to shift in the direction of the auditory perturbation (toward *sh*) following speech practice. In contrast, for the UF group, the boundary shifts toward the *s*-stimuli following speech practice, while for the PL group, passive listening of frequency-altered speech stimuli results in little change in boundary location. (b) Mean change in boundary location following the period of speech practice (AF and UF groups), or passive listening to altered stimuli (PL group). The error bars show one standard error.

thereby reducing the discrepancy between the spectrally altered *s* acoustics and the perceptual representation of the *s* category. The boundary shift for the AF group differed reliably from that of the UF group, for whom auditory feedback was not spectrally altered [AF boundary shift versus UF boundary shift: $t(18)=4.5$, $p<0.05$]. In contrast with the AF group, subjects in the UF group exhibited a shift in the /s-ʃ/ category boundary toward stimuli containing a *higher* centroid frequency (i.e., further toward the /s/-category) following the period of /s/-word practice [$t(9)=2.9$, $p<0.05$].

In order to determine whether the perceptual adaptation observed in the AF group might have been due solely to perceptual exposure to frequency-shifted /s/-words, the effects observed in the AF group were compared with those of a third group of subjects (PL group) who passively listened to a sequence of frequency altered speech stimuli—matched to the stimuli perceived by the AF group—without producing any words themselves. Compared with the relatively small boundary shift observed for the PL group, a significantly larger change in the boundary location was observed for the AF group [AF versus PL groups: $t(18)=2.84$, $p<0.05$], as shown in Figs. 4(a) and 4(b).

IV. DISCUSSION

In the present study, we have investigated speech motor adaptation to altered auditory feedback during the production of the sibilant /s/. Talkers exposed to the auditory manipulation (AF group) were found to alter their speech output to compensate for the effects of the sensory perturbation. That is, they produced the sibilant with an increased centroid frequency in order to offset the decrease in frequency in the fed-back acoustic signal. Furthermore, the compensatory change was found to persist following the sudden removal of the perturbation, indicating that the change in output was not the result of online feedback-based correction but rather reflected a change in the underlying motor plan for /s/. The finding of speech-motor learning is consistent with numerous studies of speech adaptation involving manipulations of auditory feedback (e.g., involving vowel formants and fundamental frequency—Houde and Jordan, 1998, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007) and somatosensory feedback (e.g., palatal or dental prosthesis, altered jaw path—Baum and McFarland, 1997, 2000; Jones and Munhall, 2003; Tremblay *et al.*, 2003). In comparison with the subjects in the AF condition, a group of subjects who underwent an identical set of speech production procedures without any alteration to auditory feedback (UF group) did not show a shift in speech output following the period of speech practice with /s/-initial words.

Previous studies of speech-motor learning using purely auditory manipulations of sensory feedback have examined the production of vowel formants and fundamental frequency (Houde and Jordan, 1998; Jones and Munhall, 2000; Houde and Jordan, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In contrast, the present study provides a demonstration of speech adaptation following a purely auditory perturbation of a consonantal speech sound. Studies of motor adaptation involving the production of sibilants have, of course, been carried out using perturbations that alter the shape of the vocal tract (e.g., using palatal or dental prostheses—Baum and McFarland, 1997, 2000; Jones and Munhall, 2003). The impact of these physical manipulations is complex—altering both tactile/somatosensory feedback as well as speech acoustics—therefore it is impossible to determine the extent to which subjects in these studies were compensating for changes in a particular sensory modality. In contrast, the observation of motor adaptation in the present study offers direct support for the role of auditory input in the production of the sibilant /s/.

In addition to the demonstration of speech motor adaptation in the present study, an investigation of the perceptual representation of the /s-ʃ/ contrast revealed a shift in the phoneme identification boundary following speech practice under feedback-altered conditions (AF group). Specifically, the boundary was found to shift in the same direction as the feedback manipulation (toward a lower centroid frequency, i.e., the /ʃ/ category). Such a perceptual shift is adaptive, since it has the effect of reducing the functional impact of the auditory feedback manipulation by increasing the perceptual distance between the category boundary and the subjects' altered /s/-stimuli. The extent to which this perceptual adaptation effect might have arisen solely due to the perceived change in acoustic properties of /s/-words was explored by exposing a set of similarly altered stimuli to a group of control subjects who themselves did not produce any speech (PL group). Following the exposure to frequency-shifted /s/-words (using the same testing sequence as subjects in the AF group), no overall change in these subjects' perceptual boundaries between /s/ and /ʃ/ was observed, suggesting that the perceptual changes observed in the AF group were indeed related to a perceptuomotor adaptation.

While subjects in the PL group exhibited no change in perceptual representations, subjects in the UF control group (who produced repeated /s/-words under conditions of unaltered auditory feedback) did exhibit a shift in the perceptual category boundary following the training phase, however, in this case toward a *higher* centroid frequency (i.e., in the direction of the /s/ category). Such a boundary shift is consistent with the well-known *selective adaptation effect* (Eimas and Corbit, 1973), a perceptual phenomenon in which repeated exposure to stimuli at one end of a phonetic continuum results in a shift in the phoneme identification boundary toward the repeated stimulus. Given that subjects in all three groups were exposed to repeated /s/-stimuli, the question remains as to why a selective adaptation effect was only observed in the UF group. One possibility is that the selective adaptation effect depends critically on the repeated presentation of canonical endpoint stimuli (i.e., unaltered /s/-words), as have typically been employed in studies investigating the phenomenon. Indeed, a small number of selective adaptation studies have included degraded or otherwise noncanonical speech stimuli and in spite of the subjects' tendency to categorize these stimuli as belonging to a particular phonemic category, a weaker or nonexistent selective adaptation effect was reported in comparison with end-

point stimuli (Sawusch and Pisoni, 1976; Blumstein *et al.*, 1977; Cheesman and Greenwood, 1995).

In addition to shifting perceptual category boundaries, the selective adaptation effect has been shown to impact the *production* of speech sounds in a number of studies (Cooper and Lauritsen, 1974; Cooper and Nager, 1975; Jamieson and Cheesman, 1987). Cooper and Lauritsen's (1974) seminal study demonstrated that repetitive listening to a CV syllable with an initial voiceless stop consonant (which presumably resulted in a perceptual selective adaptation effect) caused subjects to produce a shorter voice onset time (VOT) for voiceless stop consonants in CV syllables. Given this finding, it is perhaps surprising that the UF group in the present study did not show a corresponding shift in speech output. This discrepancy is likely explained by the substantially smaller number of trials in the current study as compared to that utilized by previous studies such as Cooper and Lauritsen's (1974), in which a significant shift in motor output was observed following many hundreds of trials. In addition, the impact of the selective adaptation effect on speech production has only been demonstrated in the context of a single speech parameter—VOT for voiceless plosives—with a failure to demonstrate a corresponding change in VOT output following selective adaptation to their voiced counterparts (Cooper and Lauritsen, 1974; Cooper and Nager, 1975; Jamieson and Cheesman, 1987). The apparent phonetic specificity of this perceptuomotor selective adaptation effect leaves open the question of whether a similar phenomenon would be expected in the case of fricatives.

A number of prior studies have demonstrated changes in the perception of phoneme categories under a range of listening conditions (Ladefoged and Broadbent, 1957; Miller and Liberman, 1979; Mann and Repp, 1980; Bertelson *et al.*, 2003; Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006). The question arises whether similar input processes might underlie the perceptual adaptation observed in the present study. For example, it has been shown that when listeners are exposed to phonetic stimuli that are perceptually ambiguous, e.g., a word containing a fricative that is acoustically midway between two categories (as in Norris *et al.*, 2003; Kraljic and Samuel, 2005; McQueen *et al.*, 2006), their phoneme category boundary may shift in order to reduce the ambiguity. In the present study, we investigated the possible contribution of a lexically based perceptual adaptation effect by examining the change in the /s-ʃ/ perceptual boundary in a group of subjects (PL group) who listened passively to a sequence of /s/-words that were frequency-altered in the same manner as the speech feedback presented to the AF group. The PL group showed no overall shift in their /s-ʃ/ category boundary following the perceptual exposure. The lack of a lexically based perceptual adaptation effect in the present study is likely related to the nature of the auditory stimuli involved. While previous perceptual adaptation studies have employed acoustic stimuli that were designed to be maximally acoustically and perceptually ambiguous, the /s/-productions in the present study were not degraded to such a large degree. As a result, the stimuli in the present study—while not typical /s/ tokens—remained within the /s/ category and hence were not, in fact, perceptually ambiguous.[3]

The results of the present study provide new insights into aspects of previous studies of speech adaptation that have been poorly understood, in particular, the common finding that adaptation in motor output is less than complete (Baum and McFarland, 1997; Houde and Jordan, 1998, 2002; Tremblay *et al.*, 2003; Jones and Munhall, 2005). The finding of incomplete adaptation in prior studies has been discussed primarily in terms of factors that might impact the motor planning process (although, see Houde and Jordan, 2002). These include articulatory constraints, auditory acuity, the balance of feedback and feed-forward control mechanisms, modulation of the salience of auditory feedback over time, and, in the case of purely auditory perturbations, the possible impact of unaltered somatosensory targets for speech sounds (e.g., Savariaux *et al.*, 1995; Baum and McFarland, 2000; Houde and Jordan, 2002; Purcell and Munhall, 2006a; Villacorta *et al.*, 2007). In addition, individual differences in the degree of motor adaptation may be related to speech production characteristics such as token-to-token variability and the distinctiveness of phoneme production—properties that could influence the impact of a given feedback manipulation. In contrast with these accounts, the present results suggest that speech adaptation to altered auditory feedback is not limited to the motor domain, but rather involves changes in both motor output and auditory representations of speech sounds that together act to reduce the impact of the perturbation. Thus, the contribution of both sensory and motor adaptive processes, in conjunction with those factors already suggested, might offer a better account of speech adaptation to altered sensory conditions.

These results appear in keeping with the long-standing proposal that speech perception and speech production are closely linked processes, as first detailed in the motor theory of speech perception (Liberman *et al.*, 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000; see also Galantucci *et al.*, 2006), itself an extension of earlier theories of perception in which motor actions were viewed as possible integral components of the perceptual process (Berkeley, 1709; Washburn, 1926; Festinger *et al.*, 1967). The findings also appear consistent with a growing number of behavioral, neuroimaging, and neurophysiological studies providing supporting evidence for sensorimotor interactions during both speech perception and production (Cooper and Lauritsen, 1974; Fadiga *et al.*, 2002; Watkins *et al.*, 2003; Wilson *et al.*, 2004; Sams *et al.*, 2005; Pulvermuller *et al.*, 2006; Gentilucci and Bernardis, 2007; Meister *et al.*, 2007; Skipper *et al.*, 2007; Tourville *et al.*, 2008). The results of the present study provide new behavioral evidence showing not only a link between the processes underlying speech perception and production, but a functional and plastic change involving both input and output processes simultaneously.

ACKNOWLEDGMENTS

Supported by research grants from NSERC-Canada, a postdoctoral fellowship from the Centre for Research on

Language, Mind and Brain, McGill University, Canada (D.M.S.), and a Richard H. Tomlinson postdoctoral fellowship from McGill University, Canada (M.S.).

[1]Importantly, the AF and UF groups did not differ appreciably in /s/ duration (584 ms versus 613 ms). In addition, for the AF group, the mean fricative duration was very similar prior to and following the period of training under conditions of altered auditory feedback (584 ms versus 593 ms), which indicates that differences in duration did not play a role in the observed adaptation effects.

[2]As described in Sec. III, vowel context was not found to have a reliable impact on the magnitude of the /s/ motor adaptation effect in the AF group.

[3]In order to verify that the three-semitone frequency shift did not result in ambiguous stimuli, a small control study was carried out as follows: Single tokens of the words "sue" and "shoe" were selected from the baseline productions of the first five subjects in the AF group. The five tokens of "sue" were then played back through the DSP, applying a frequency shift of −3.0 semitones, and the output was digitally recorded. All speech stimuli (unshifted "sue" and "shoe," plus the frequency shifted "sue") were normalized in order to remove any differences in fricative amplitude and duration within each of the five speakers. Vowel differences were also eliminated by concatenating the normalized fricatives with a single vowel (taken from the unfiltered production of "sue") for each subject, resulting in three stimuli per talker that differed solely in the frequency composition of the fricatives. A perceptual test was carried out in which 6 female listeners made phoneme identity judgments ("s" versus "sh") on 85 speech stimuli played back through headphones under conditions comparable to those used for perceptual testing in the main study. The stimulus set contained five repetitions of the 15 different speech stimuli presented in randomized order (totaling 25 repetitions of each fricative), plus 10 initial practice trials (randomly selected). As expected, all eight listeners consistently labeled the unaltered /s/-stimuli as "s" (averaging 99% "s" responses across listeners) and the unaltered /ʃ/-stimuli as "sh" (2% "s" responses). Notably, subjects were also highly consistent in their labeling of the frequency-shifted /s/-stimuli as "s" (99% "s" responses). Thus, while the /s/-productions were altered by the DSP, the magnitude of the frequency shift was not great enough to render the stimuli phonemically ambiguous.

Abbs, J. H., and Gracco, V. L. (**1984**). "Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech," J. Neurophysiol. **51**, 705–723.

Baum, S. R., and McFarland, D. H. (**1997**). "The development of speech adaptation to an artificial palate," J. Acoust. Soc. Am. **102**, 2353–2359.

Baum, S. R., and McFarland, D. H., (**2000**). "Individual differences in speech adaptation to an artificial palate," J. Acoust. Soc. Am. **107**, 3572–3575.

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.

Berkeley, G., (**1709**). "An essay towards a new theory of vision," in *The Works of George Berkeley*, edited by A. A. Luce and T. E. Jessop (Nelson, London).

Bertelson, P., Vroomen, J., and De Gelder, B., (**2003**). "Visual recalibration of auditory speech identification: A McGurk aftereffect," Psychol. Sci. **14**, 592–597.

Blumstein, S. E., Stevens, K. N., and Nigro, G. N. (**1977**). "Property detectors for bursts and transitions in speech perception," J. Acoust. Soc. Am. **61**, 1301–1313.

Bradlow, A. R., and Bent, T., (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Cheesman, M. F., and Greenwood, K. G. (**1995**). "Selective adaptation by context-conditioned fricatives," J. Acoust. Soc. Am. **97**, 531–538.

Clarke, C. M., and Garrett, M. F., (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cooper, W. E., and Lauritsen, M. R. (**1974**). "Feature processing in the perception and production of speech," Nature (London) **252**, 121–123.

Cooper, W. E., and Nager, R. M. (**1975**). "Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model," J. Acoust. Soc. Am. **58**, 256–266.

Eimas, P. D., and Corbit, J. D. (**1973**). "Selective adaptation of linguistic feature detectors," Cogn. Psychol. **4**, 99–109.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G., (**2002**). "Speech listening specifically modulates the excitability of tongue muscles: A TMS study," Eur. J. Neurosci. **15**, 399–402.

Festinger, L., Burnham, C. A., Ono, H., and Bamber, D. (**1967**). "Efference and the conscious experience of perception," J. Exp. Psychol. **74**, 1–36.

Galantucci, B., Fowler, C. A., and Turvey, M. T., (**2006**). "The motor theory of speech perception reviewed," Psychon. Bull. Rev. **13**, 361–377.

Gentilucci, M., and Bernardis, P., (**2007**). "Imitation during phoneme production," Neuropsychologia **45**, 608–615.

Gracco, V. L., and Abbs, J. H. (**1985**). "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes," J. Neurophysiol. **54**, 418–432.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disord. **39**, 350–365.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I., (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jamieson, D. G., and Cheesman, M. F. (**1987**). "The adaptation of produced voice-onset time," J. Phonetics **15**, 15–27.

Jones, J. A., and Munhall, K. G., (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Jones, J. A., and Munhall, K. G., (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

Jones, J. A., and Munhall, K. G., (**2005**). "Remapping auditory-motor representations in voice production," Curr. Biol. **15**, 1768–1772.

Jongman, A., Wayland, R., and Wong, S., (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kawahara, H. (**1995**). "Hearing voice: Transformed auditory feedback effects on voice pitch control," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'95) Workshop on Computational Auditory Scene Analysis, pp. 143–148.

Kawato, M. (**1999**). "Internal models for motor control and trajectory planning," Curr. Opin. Neurobiol. **9**, 718–727.

Kraljic, T., and Samuel, A. G., (**2005**). "Perceptual learning for speech: Is there a return to normal?," Cognit Psychol. **51**, 141–178.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lane, H., Denny, M., Guenther, F. H., Hanson, H. M., Marrone, N., Matthies, M. L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M., (**2007**). "On the structure of phoneme categories in listeners with cochlear implants," J. Speech Lang. Hear. Res. **50**, 2–14.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Liberman, A. M., and Whalen, D. H., (**2000**). "On the relation of speech to language," Trends Cogn. Sci. **4**, 187–196.

Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [zh]-[s] distinction," Percept. Psychophys. **28**, 213–228.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

McQueen, J. M., Norris, D., and Cutler, A., (**2006**). "The dynamic nature of speech perception," Lang Speech **49**, 101–112.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M., (**2007**). "The essential role of premotor cortex in speech perception," Curr. Biol. **17**, 1692–1696.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Nasir, S. M., and Ostry, D. J., (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**, 1918–1923.

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Norris, D., McQueen, J. M., and Cutler, A., (**2003**). "Perceptual learning in speech," Cognit Psychol. **47**, 204–238.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-

Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**, 227–250.
Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y., (**2006**). "Motor cortex maps articulatory features of speech sounds," Proc. Natl. Acad. Sci. U.S.A. **103**, 7865–7870.
Purcell, D. W., and Munhall, K. G., (**2006a**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.
Purcell, D. W., and Munhall, K. G., (**2006b**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Sams, M., Mottonen, R., and Sihvonen, T., (**2005**). "Seeing and hearing others and oneself talk," Brain Res. Cognit. Brain Res. **23**, 429–435.
Savariaux, C., Perrier, P., and Orliaquet, J. P. (**1995**). "Compensation strategies for the perturbation of the rounded," J. Acoust. Soc. Am. **98**, 2428–2422.
Sawusch, J. R., and Pisoni, D. B. (**1976**). "Simple and contingent adaptation effects for place of articulation in stop consonants," Percept. Psychophys. **23**, 125–131.
Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L., (**2007**). "Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception," Cereb. Cortex **17**, 2387–2399.
Tourville, J. A., Reilly, K. J., and Guenther, F. H., (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–1443.
Tremblay, S., Shiller, D. M., and Ostry, D. J., (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.
Villacorta, V. M., Perkell, J. S., and Guenther, F. H., (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.
Washburn, M. F. (**1926**). "Gestalt psychology and motor psychology," Am. J. Psychol. **37**, 516–520.
Watkins, K. E., Strafella, A. P., and Paus, T., (**2003**). "Seeing and hearing speech excites the motor system involved in speech production," Neuropsychologia **41**, 989–994.
Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M., (**2004**). "Listening to speech activates motor areas involved in speech production," Nat. Neurosci. **7**, 701–702.