# An investigation of the relation between sibilant production and somatosensory and auditory acuity

Satrajit S. Ghosh[a)]
*Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

Melanie L. Matthies
*Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215*

Edwin Maas
*Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721*

Alexandra Hanson
*Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts 02215*

Mark Tiede
*Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511*

Lucie Ménard
*Département de linguistique et de didactique des langues, Université du Québec à Montréal, Montréal H3C 3P8, Canada*

Frank H. Guenther
*Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215*

Harlan Lane
*Department of Psychology, Northeastern University, Boston, Massachusetts 02115*

Joseph S. Perkell
*Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

(Received 30 April 2009; revised 19 July 2010; accepted 27 August 2010)

The relation between auditory acuity, somatosensory acuity and the magnitude of produced sibilant contrast was investigated with data from 18 participants. To measure auditory acuity, stimuli from a synthetic sibilant continuum ([*s*]-[ʃ]) were used in a four-interval, two-alternative forced choice adaptive-staircase discrimination task. To measure somatosensory acuity, small plastic domes with grooves of different spacing were pressed against each participant's tongue tip and the participant was asked to identify one of four possible orientations of the grooves. Sibilant contrast magnitudes were estimated from productions of the words 'said,' 'shed,' 'sid,' and 'shid'. Multiple linear regression revealed a significant relation indicating that a combination of somatosensory and auditory acuity measures predicts produced acoustic contrast. When the participants were divided into high- and low-acuity groups based on their median somatosensory and auditory acuity measures, separate ANOVA analyses with sibilant contrast as the dependent variable yielded a significant main effect for each acuity group. These results provide evidence that sibilant productions have auditory as well as somatosensory goals and are consistent with prior results and the theoretical framework underlying the DIVA model of speech production.
 [DOI: 10.1121/1.3493430]

PACS number(s): 43.70.Mn, 43.70.Bk, 43.71.An, 43.70.Jt [PEI] Pages: 3079–3087

## I. INTRODUCTION

In speech communication an important goal of any speaker is to maximize intelligibility. One way in which speakers achieve this goal is to produce speech sound contrasts that are as large as necessary and possible. However, speakers vary in their ability to produce such contrasts. Prior work (e.g., Perkell *et al.*, 2004a, 2004b) from our laboratory has shown that production differences among speakers in certain vowel and sibilant contrasts are related to their auditory acuity. Speakers with greater acuity for vowel and sibilant contrasts tended to produce greater contrasts. The purpose of this study is to explore the effects of auditory and somatosensory acuity on the sibilant contrast ([s-ʃ]) as produced by speakers of American English. Examining such re-

[a)]Author to whom correspondence should be addressed. Electronic mail: satra@mit.edu

lations between perception and production should also provide insight into a central question about speech motor control and speech acquisition: what are the motor control parameters (internal representations or targets) that underlie the production of different speech sounds?

The current investigation was guided by the theoretical framework underlying the DIVA model (Guenther, 1994, 1995; Guenther *et al.*, 1998, 2006), which postulates that speech sounds have associated somatosensory and auditory goal regions or targets. According to the model, accurately producing a speech sound requires controlled vocal tract movements that generate patterns of somatosensory and auditory feedback that are within specific sensory target regions that are associated with the speech sound. Several lines of evidence support the idea that speech sounds have distinct auditory and somatosensory goals.

Studies of motor equivalence (e.g., Perkell *et al.*, 1993; Hughes and Abbs, 1976) have demonstrated that somewhat different vocal tract configurations can be used to reach consistent acoustic (also referred to here as "auditory") goals. The role of auditory goals is also addressed by auditory perturbation studies that indicate that changes are made in the programming of speech movements when speakers' auditory feedback is perturbed. In those studies, compensatory responses were observed in response to perturbations of fundamental frequency (Jones and Munhall, 2000, 2005), the first formant of vowels (Houde and Jordan, 1998; Purcell and Munhall, 2006; Villacorta *et al.*, 2007; Tourville *et al.*, 2008; Cai *et al.*, 2008) and more recently in response to perturbation of the sibilant [s] (Shiller *et al.*, 2009). Further evidence of auditory goals comes from studies investigating the effects of changes in the state of speakers' hearing. For example, produced speech sound contrasts are reduced in normal-hearing participants when auditory feedback is suppressed using masking noise (Perkell *et al.*, 2007). Postlingually deafened cochlear implant candidates, who have little or no auditory feedback, show reduced contrasts prior to implantation. When auditory feedback was made available by implantation, implant users increased their produced vowel (Vick *et al.*, 2001; Lane *et al.*, 2007) and sibilant contrasts (Matthies *et al.*, 1994; Lane *et al.*, 2007), reduced the variability in their productions of allophones of [ɹ] (Matthies *et al.*, 2008) and increased the contrast between [ɹ] and [l] (Perkell *et al.*, 2001).

Results from studies using mechanical perturbation during speech (e.g., Gracco and Abbs, 1989; McFarland *et al.*, 1996; Nasir and Ostry, 2008) provide evidence about the role of somatosensory goals. Jones and Munhall (2003) reported adaptation to a dental prosthesis that altered the production of the sibilant [s]. They noted that compensation (measured acoustically) was not impeded by noise-masked auditory feedback. However, listeners perceived the quality of the productions to be closer to the speakers' unperturbed production when the speakers' auditory feedback was not masked. These results indicate a reliance on both somatosensory and auditory cues for [s] production. It is well known that postlingually deafened individuals are often able to maintain intelligible speech for a long period of time. This may be due, at least in part, to the continued availability of somatosensory feedback, which presumably remain largely intact even after decades of profound deafness. Nasir and Ostry (2008) perturbed jaw movements in speakers who used cochlear implants. The participants compensated for the perturbation even when their implants were turned off, indicating their reliance on somatosensory feedback. A study by Niemi *et al.* (2006) provides further evidence of the role of somatosensory feedback in speech production. They reported changes in sibilant spectra when orosensory feedback was reduced using local anesthesia of the right lingual nerve. It can be inferred from such observations that somatosensory goals play an important role in maintaining programmed speech movements in the absence of auditory feedback. Some further support for the use of somatosensory goals was provided by Wohlert and Smith (1998), who reported that older participants, who had low labial tactile acuity showed greater variability in their speech movements when compared to younger participants, who had higher tactile acuity.

Results from these and other studies support a role for auditory and somatosensory goals in programming of speech movements, which leads to the inference that sensory feedback is of central importance for speech motor planning. Since sensory perception seems to play such an important role in speech production, our theoretical framework posits that some of the widely-observed inter-speaker production variability is related to differences in their perceptual capacities. The framework posits that speakers with higher perceptual acuity will produce speech sounds with greater contrast than speakers with lower acuity (cf., Perkell, 2009, 2010).

Several studies have shown cross-speaker relations between perception and production of speech sounds(see Perkell *et al.*, 2004a, for a discussion). Perkell *et al.* (2004a) reported that the magnitude of produced vowel contrasts was greater among participants who had better acuity in discriminating between the same vowels. Villacorta *et al.* (2007) reported that the amount of speakers' compensatory adaptation to F1 shifts in their auditory feedback was correlated with their auditory acuity. In a study investigating the production of the sibilants [s] and [ʃ], Perkell *et al.* (2004b) reported that speakers with higher auditory acuity for the sibilant contrast produced greater acoustic contrast between the sibilants. They suggested that in addition to an acoustic goal, a possible somatosensory goal for the sibilant [s] is contact between the underside of the tongue tip with the lower alveolar ridge. It was hypothesized that speakers who supplemented auditory feedback with somatosensory feedback from contact of the tongue tip with the alveolar ridge, would produce [s] tokens acoustically more distinct from [ʃ]. They tested whether such contact was made on each of a number of trials in which subjects pronounced words containing [s] and [ʃ]. The results showed that speakers with a larger mean proportion of [s] productions made with contact and [ʃ] productions with no contact had greater acoustic contrast than speakers who less consistently showed this difference in contact between the two sounds. Although the results were suggestive of a role for somatosensory information in sibilant productions, the study had two methodological shortcomings: (i) it

did not fully explore the issue by measuring somatosensory acuity; and (ii) the auditory acuity measure was not fine grained.

The sibilants [s] and [ʃ] can be thought of as having prominent somatosensory and auditory goals (Perkell *et al.*, 2000, 2004b; Perkell, 2009, 2010). In the auditory domain, they are distinguished by measures of the distribution of energy in their noise spectra such as spectral mean, skew and kurtosis (Forrest *et al.*, 1988). These differences reflect the fact that the centroid of the spectrum of [s] is higher than that of [ʃ]. In the somatosensory domain, they are produced with differences in tongue-blade shape and position within the oral cavity and therefore will differ in produced patterns of proprioceptive and tactile feedback. As suggested above, one such tactile difference is the contact between the tongue and the lower alveolar ridge. Thus the sibilants are ideally suited for exploring the hypothesis that *contrast distance will be positively correlated, across speakers, to both somatosensory and auditory acuity*. In order to test this hypothesis, a set of experiments was conducted to measure produced sibilant acoustic contrast, somatosensory acuity and auditory acuity, using the same group of participants for all three experiments. Experimental methods are provided next.

## II. METHODS

### A. Data collection and feature extraction

A group of 18 young adult speakers of American English (10 females, 8 males) participated in three separate experimental sessions. Data were acquired to estimate three measures of interest for each participant: (i) somatosensory acuity of the tongue tip; (ii) auditory acuity for the sibilant contrast; and (iii) difference between measures of the energy distribution in the produced spectra of [s] and [ʃ], called "contrast distance." All experimental procedures were approved by the Institutional Review Board at Massachusetts Institute of Technology. The details of methods used are provided in the following paragraphs.

#### *1. Session 1: Measurement of somatosensory acuity*

Prior studies of oral somatosensory sensitivity or acuity have typically used a two-point discrimination task (e.g., Ringel and Ewanowski, 1965; McNutt, 1975, 1977; Maeyama and Plattig, 1989; Sato *et al.*, 1999). According to the results of these studies, two-point discrimination thresholds on the tongue are in the 1.5–2.5 mm range. However, according to Van Boven and Johnson (1994), "the conventional test, the two-point discrimination task, does not measure the limit of spatial resolution and it yields variable results because it does not control nonspatial cues." To overcome the limitations of the two-point discrimination task, Van Boven and Johnson (1994) used a grating orientation discrimination test to determine the limits of spatial resolution at the lip, tongue and finger. Wohlert (1996) and Wohlert and Smith (1998) used a similar grating orientation discrimination task in studies that related tactile acuity of the lips to variability in speech production.

The production of sibilants requires precise positioning and shaping of the tongue to create a narrow constriction between the tongue blade and anterior palate and a cavity anterior to the constriction with resonances that, when excited by turbulence noise generated at the constriction, will produce a sound with specific spectral properties (Stevens, 1998). Tactile acuity of the tongue tip is hypothesized to have an important influence on the produced resonator configuration, particularly for sensing contact of the tongue with the lower incisors. Since such contact was hypothesized to be a somatosensory goal for [s] but not for [ʃ] (Perkell *et al.*, 2004b), we measured tactile acuity of the tongue tip for the current experiment.

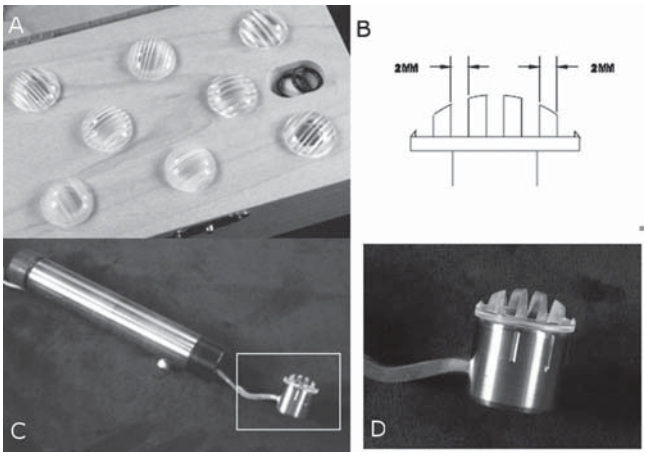


FIG. 1. JVP Domes and custom probe. (a) The set of 8 domes used in the study. (b) Grid spacing on one dome. (c) Custom holder used to apply pressure. The strain gauges are inside the handle and are not visible. (d) Magnified image of the region marked by the white box in (c).

Considering the shortcomings of two-point discrimination tests and our need for a relatively simple way to quantify differences in somatosensory acuity across participants, we employed a grating-orientation judgment task using JVP Domes™ [JVPD; Fig. 1(a)]. According to Van Boven and Johnson (1994), JVPD provide one of the most sensitive approaches among readily available techniques for measurement of somatosensory acuity. The JVPD are a set of 8 plastic domes, each 19 mm in diameter, containing equidistant bars and grooves with different widths: 0.35 (#1), 0.50, 0.75, 1.00, 1.25, 1.50, 2.00 and 3.00 (#8) mm. Figure 1(b) shows a cross-section of the dome with 2.00 mm bar and groove widths. To press the domes against a subject's protruded tongue tip with a relatively consistent force and duration, we built a custom holder [Fig. 1(c)] with a handle and a cylindrical receptacle into which one of the domes could be inserted and set at one of four distinct orientations. The receptacle was mounted on the end of a strain-gauge cantilever beam contained inside the handle. The strain gauges were connected to a bridge amplifier, which also low-pass filtered the signal at 500 Hz. A National Instruments A/D device sampled the output of the bridge amplifier at 1000 Hz. A 2N weight was used to calibrate the sampled signal from the strain gauge. A MATLAB (The Mathworks, Natick, MA) script was developed to control the experiment. A near real-time algorithm monitored the applied force level to aid the experimenter in maintaining the pressure within a consistent range for a consistent duration. When the force reached a

specified level, a beep was emitted over headphones, indicating to the experimenter that the current force should be maintained. Then a second beep occurred after 0.5 s, indicating that the dome should be withdrawn. The experimental setup did not support continuous recording of the values of the applied force.

*a. Experimental procedure* The measurement of somatosensory acuity consisted of two parts: one that probed the anterior palate and the other, mucosa of the tongue tip. With the dome grooves oriented at differing angles with respect to the mid-sagittal plane, the experimenter used the holder to press the dome against the palate or the tongue tip. Participants were asked to identify the orientation of the grooves, by using a mouse to click on one of the response options presented in front of them on a computer screen.

In the first part, the sensitivity of the anterior palate was probed, since the production of the sibilants involves formation of a narrow groove in the tongue blade as it is pressed against the palate in this region. Preliminary testing revealed that speakers' perception of grid orientation with the hard palate was very poor. Therefore, this test used only the coarsest grating (#8, 3 mm) and only two orientations. Using the holder, the dome was pressed against the palate with between 0.2 N and 0.6 N of force for approximately 500 ms. In any given trial, the grooves of the dome were oriented either vertically or horizontally with respect to the midline of the palate, comprising a 2-alternative forced choice task.

In the second part, the sensitivity of the mucosa of the tongue tip was probed with all 8 domes. Using the holder, the dome was pressed against the tip of the subject's protruded tongue with a force between 0.002 N and 0.5 N for approximately 500 ms. The grooves were oriented in four possible directions (horizontal, vertical, left diagonal, right diagonal) with respect to the midline of the tongue (a 4-alternative forced choice paradigm).

At the beginning of each part, 10 practice trials with different orientations were conducted using the #8 probe in order to familiarize the participants with the procedure. The test phase followed the practice trials. For each part, the test phase was divided into 3 blocks, with each of the consecutive blocks containing 3, 4 and 3 repetitions respectively of each orientation of every groove-width used in that part. The order of groove orientations was randomized. During the course of each part, a participant experienced 10 repetitions of the same orientation from any given grating.

Throughout the experiment, participants were provided visual feedback about whether their choice was correct at the end of each trial. The participants also wore goggles that prevented them from viewing the probe and inferring the orientation of the grating. The data acquisition and experimental protocol were controlled using custom MATLAB scripts.

*b. Determining somatosensory acuity* Two measures of somatosensory acuity (SA) for each dome were estimated from the data from every participant. These were (i) percent correct; and (ii) a standardized measure that took response bias into account. The "somatosensory acuity percent correct" (SAPC) was computed as the number of hits (correct choice of orientation) divided by the total number of trials for a given dome. The standardized measure was computed for a given dome and a specific orientation (e.g., vertical) by computing the z-score of the difference between hits and false alarms (e.g., selecting vertical orientation although a different orientation was presented). The mean of these estimates across orientations was found for each dome. SAPC and the standardized measure were highly correlated across grid sizes and participants ($r=0.98$, $p<0.001$, $n=144$). Each participant's maximum SAPC was used for subsequent analysis regardless of which dome the maximum came from. This measure was also correlated with mean percentage correct ($r=0.91$), percentage correct at probe #6 ($r=0.79$), percentage correct at probe #3 ($r=0.93$) and slope of a line fit to the percentage correct data ($r=0.89$).

### 2. Session 2: Measurement of auditory acuity

In the second session, the participants performed labeling and discrimination tests aimed at measuring their auditory acuity along a [s-ʃ] continuum. The Klatt synthesizer (Klatt and Klatt, 1990) was used to generate a sibilant continuum in 841 steps for the continuum "said"-"shed" (Klatt parameter documents for the end points are available as supplementary material). The synthesis parameters were derived from natural utterances of "said" and "shed" spoken by a male speaker. Each token was analyzed to extract segment boundaries corresponding to consonant and vowel portions of the utterance. Formant frequency, fundamental frequency and energy contours were extracted algorithmically for each segment. The frequency values of the spectral peaks were estimated from a mean spectral representation of the entire sibilant segment from each end point ("said:" 808, 2032, 3917, 4892, 5294, 6304; "shed:" 1211, 2272, 2957, 3658, 4741, 5830; in Hz). These were used as formant values for the cascade branch of the synthesizer and excited with a fricative noise source that changed in amplitude (corresponding to the energy contour of the sibilant segment). The continuum was created by morphing between the sibilant segment parameters from the two end points while holding the vowel and the final consonant segment parameters constant. The fundamental frequency of every utterance was scaled to have a mean value of 165 Hz, a value that falls between a prototypical male and a prototypical female voice. Informal perception tests indicated that the synthesized tokens sounded natural.

*a. Experimental procedure* The study of auditory acuity consisted of two parts. In the first part, each participant's category boundary was established using a labeling task. Participants heard tokens that were selected from 11 equally spaced intervals between the end points of the [s-ʃ] continuum. Ten repetitions of each of the 11 tokens were presented with random ordering. Participants were asked to label each token as 'said' or 'shed'. Logistic functions were fit to each category of the labeling data and the frequency value at the intersection of the two functions were used to determine each participant's boundary location for the continuum.

In the second part, participants performed a spectral discrimination test around the boundary determined in the first part. The discrimination test used a 4-interval, 2-alternative forced choice (4I-2AFC) task, in which participants heard

the sequence A-B-A-A or A-A-B-A and had to select whether the 2nd or 3rd item was different from the rest. Participants were provided feedback about whether their response was correct. An adaptive staircase procedure was used in which the initial separation between A and B was large. After each trial, the separation decreased by 1 step following a correct response and increased by 3 steps following an incorrect response (Kaernbach, 1991). At the end of each trial, the size of the step by which the increase or decrease took place was set to 10% of the separation between the stimuli used in that trial. Thus, as the separation became smaller, the step size reduced proportionately. The staircase was terminated after 14 reversals or 80 trials. A *separation index* (SI) for the staircase was estimated as the average of the separation between the stimuli A and B at each of the final four reversals. Each participant completed four such staircase runs.

*b. Determining auditory acuity* For each of the four runs, an auditory JND (just noticeable difference) was calculated as the value corresponding to the difference in spectral mean (in Hz) of two synthetic stimuli separated by the SI around the category boundary. The JND value from each participant's final run was used for the statistical analysis.

### 3. Session 3: Measurement of sibilant contrast distance

In this session, each participant's productions of the words 'said', 'shed', 'sid' and 'shid' were recorded. The participant was seated in a sound-attenuating room and the acoustic signal was transduced by a unidirectional microphone positioned about 14 in. from the participant's lips. The microphone output was sampled at 60 kHz, low-pass filtered at 10 kHz and downsampled to 20 kHz.

*a. Experimental procedure* Participants produced 15 repetitions of each word embedded in a carrier phrase ("Say… for us") and were spoken in 'Clear', 'Casual' and 'Fast' conditions. In the 'Clear' condition, participants were asked to speak as if "they were talking to somebody who had difficulty understanding English." In the 'Casual' condition, they were asked to speak as if "they were talking to a friend" and in the 'Fast' speaking condition, participants was asked to speak rapidly as if "they were late for an appointment." Each block of productions comprised 15 repetitions of each word in each speaking condition. The order of the trials within a block and the order of speaking conditions across blocks were both block-randomized.

*b. Determining contrast distance* Spectra were estimated in the middle of the sibilant, which was generally at the time of the RMS peak of the sound pressure signal during the sibilant from each utterance, and spectral moments (mean, skewness and kurtosis) were calculated using a custom MATLAB implementation of the methods of Forrest *et al.* (1988). The spectral moments of each participant's production were standardized by the mean and standard deviation of the appropriate gender group. The contrast distance (CD) for each participant in each speaking condition was calculated as the average Euclidean difference between [s] and [ʃ] in 3-D space defined by the standardized spectral moments.

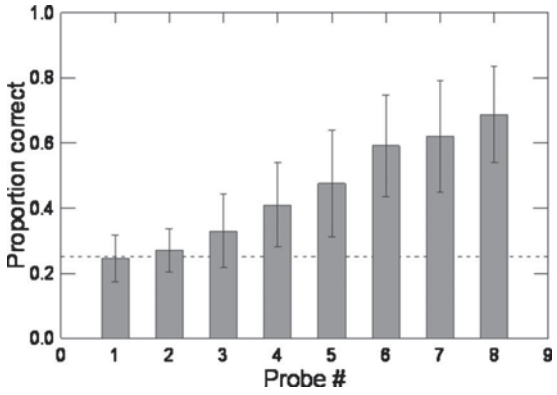


FIG. 2. Mean and standard deviation of the proportion correct for each probe averaged across all participants. Chance (0.25) performance (dotted line) was observed for probes with finest grids (#1, #2).

## III. RESULTS

Participants were unable to identify the orientation of the grating when the probe was applied to the anterior palate at better than chance levels. Thus, these measurements were not included in further analyses. The results presented below refer only to the tactile sensitivity data derived from the testing on the mucosa of the tongue tip.

Figure 2 plots proportion correct identification for each probe averaged across all participants. Chance performance (0.25) was obtained with the probes with the finest grids (#1, #2). The figure shows monotonically increasing performance with increasingly larger grid spacing. Figure 3 includes the distribution of mean proportion correct for each probe and each participant. All participants performed consistently above chance by dome #6. The finding of a consistent pattern of increasing proportion correct with increasing groove width across participants indicates that any small, uncontrolled variations of pressure applied by the experimenter had a negligible influence on the acuity measure. The maximum percent correct is represented by the topmost data-point (open circle) in each column. These values, which show substantial cross-participant variation, were used in the correlations with the participants' measures of acoustic contrast distance and auditory acuity. No participant achieved more than 90% correct, even on the coarsest grating. As mentioned above, previous studies using two-point discrimination have reported somatosensory acuity in the range of 1.5–2 mm (e.g., Ringel and Ewanowski, 1965; McNutt, 1975, 1977; Maeyama and Plattig, 1989; Sato *et al.*, 1999). The results from the grating orientation task indicate that the task used in the current study can elicit above chance performance in groove widths as low as 0.75 mm (probe #3).

Figure 4 shows that individuals with the highest contrast distance tended to have some of the highest scores for somatosensory acuity (high SAPC) and auditory acuity (low JND). To quantify this relationship and to test the hypothesis, contrast distance (CD) was correlated separately with the best somatosensory percent correct (SAPC) and with the auditory JND. These correlations, plotted in the left two panels of Fig. 4, were statistically significant at $p<0.05$ (CD vs. SAPC: $r=0.44$, $p=0.034$; CD vs. JND: $r=-0.41$, $p=0.044$). A multiple linear regression was statistically significant

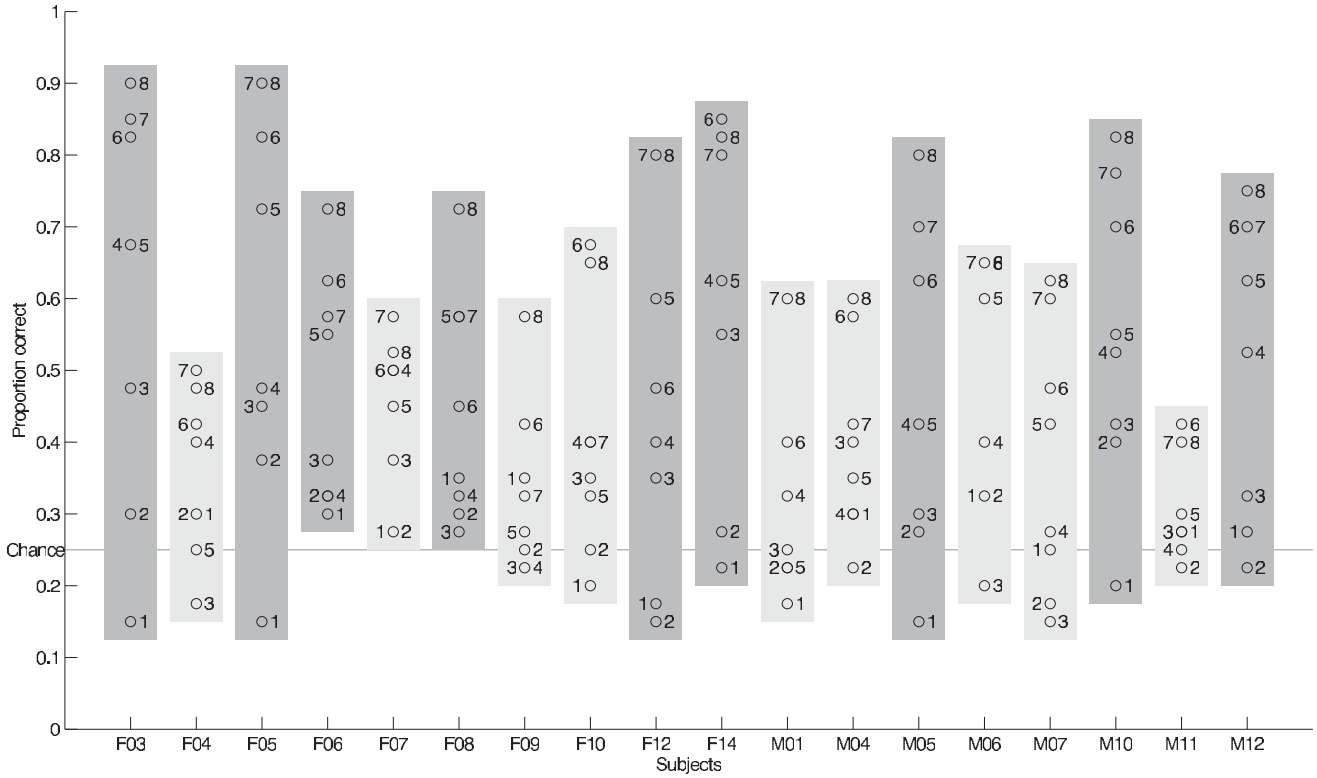


FIG. 3. Proportion correct for every participant and probe (identified by numbers). Most participants had difficulty with probes 1 and 2. All participants were above chance by probe #6. Somatosensory acuity (SAPC) was defined as the best performance (highest proportion correct) across probes for each participant (top open circle in each column). Median split groups are shown in different shades (lighter shade—lower acuity, darker shade—higher acuity).

(CD=3.9*SAPC−.002*JND+0.66: $r^2=0.43$, $p<0.02$), indicating that a combination of somatosensory and auditory acuity measures predicts produced acoustic contrast. Nonparametric Spearman's rho-based partial correlations between CD and SAPC or JND when controlling for the other acuity measure (JND or SAPC) were also significant (CD vs SAPC: $r=0.43$ $p<0.04$; CD vs JND: $r=-0.46$ $p<0.03$). (Since the hypothesis specified a positive correlation between contrast distance and acuity, one-tailed tests were used above.) A correlation between SAPC and JND, shown in the right panel of Fig. 4, was not significant (SAPC vs JND: $r=0.15$ $p<0.56$, two-tailed).

In the prior study of sibilant production and perception (Perkell *et al.*, 2004b), the relatively insensitive test of auditory acuity resulted in ceiling performance for a number of participants. For this reason, the correlation analyses were based on median splits of the data, according to participants' acuity. In order to compare the current results with the earlier ones, the current data were treated in the same way, dividing the participants into low- and high-acuity groups, based on median splits of their scores for somatosensory and auditory acuity. In an ANOVA, the effect of Auditory group (above or below the median) on contrast distance was significant $F(1,268)=41.686$ ($p<0.001$), the effect of vowel was not

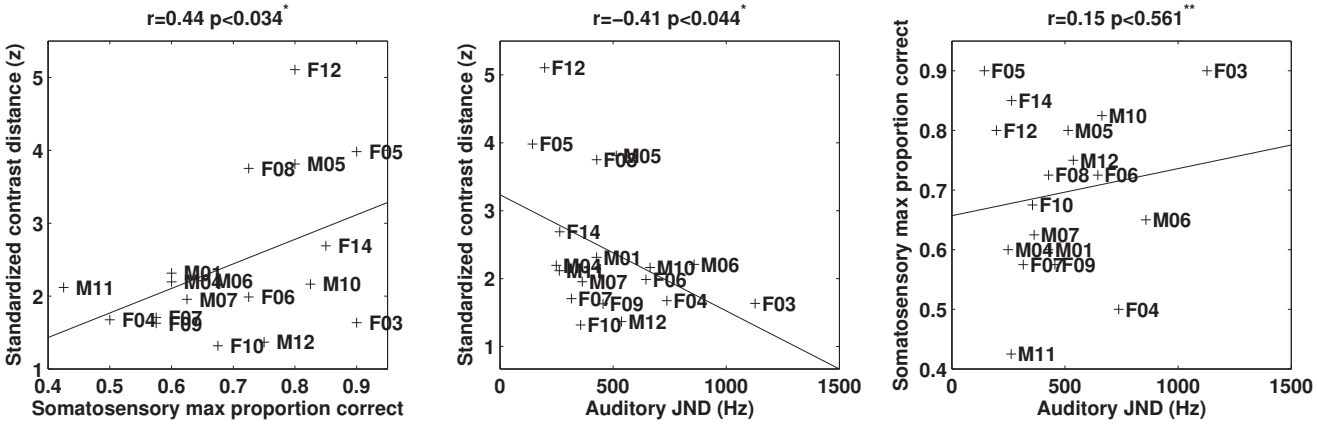


FIG. 4. Correlations between standardized contrast distance and somatosensory maximum proportion correct (left), between standardized contrast distance and auditory JND (middle) and between somatosensory maximum proportion correct and auditory JND (right). ( *-one-tailed, **-two-tailed).

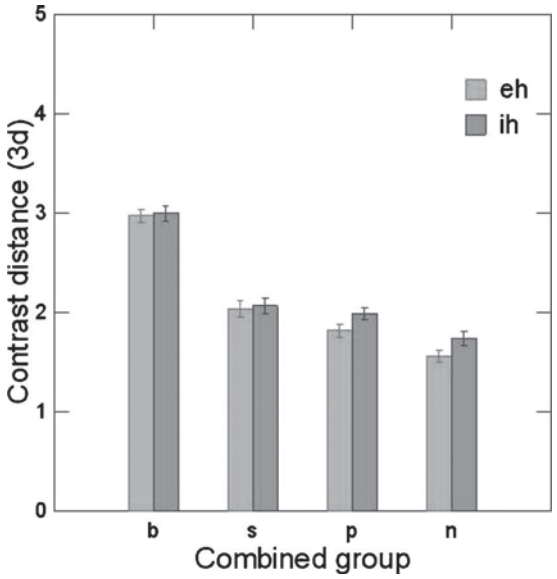


FIG. 5. Differences in 3-D contrast distance for said/shed and sid/shid as a function of group based on acuity and vowels ('eh' as in said, 'ih' as in sid; Groups-b: SomGrp=1, AudGrp=1; s: SomGrp=1, AudGrp=0; p: SomGrp=0, AudGrp=1; n: SomGrp=0, AudGrp=0; 1 indicates higher than median, and 0 indicates lower than median).

significant, speaking condition was significant $F(2,536)=30.183$ ($p<0.001$) and the Group x Condition interaction was also significant $F(2,536)=5.405$ ($p<0.005$). Other interactions were not significant. In another ANOVA, the effect of the Somatosensory group (above or below the median) on contrast distance was significant $F(1,268)=84.608$ ($p<0.001$), vowel $F(1,268)=11.069$ ($p<0.001$), condition $F(2,536)=30.891$ ($p<0.001$) and the Group $\times$ Condition interaction $F(2,536)=11.819$ ($p<0.001$) were all significant. Figure 5 shows the standardized contrast distance as a function of joint group membership. The contrast distance decreases as the acuity decreases. This is true for both vowels when pooled across the different speaking rates.

## IV. DISCUSSION

The primary aim of this study was to investigate the hypothesis that the magnitudes of produced acoustic contrast between the sibilants ([s],[ʃ]) are positively correlated with measures of the participants' somatosensory and auditory acuity. The results of the three experiments show cross-speaker variation in production of the acoustic contrast between the sibilants [s] and [ʃ] as well as in somatosensory acuity of the tongue tip and auditory acuity for the same contrast. Both auditory and somatosensory acuity yielded a statistically significant positive cross-speaker correlation with produced sibilant contrast, thereby supporting the hypothesis. In addition, the combination of somatosensory and auditory acuity yielded a statistically significant prediction of participants' produced sibilant contrast pooled across two vowel contexts and three different speaking rates, also supporting the hypothesis. When the participants were separated into high and low acuity groups based separately on their measures of somatosensory and auditory acuity, two separate ANOVA showed significant main effect of group for each kind of acuity. It is important to note that somatosensory acuity (SAPC) and auditory acuity (JND) were not correlated, showing that each type of acuity makes unique contributions to the produced contrast distance for sibilants. The results from a non-parametric, Spearman's rho-based partial correlation indicate that each of these acuity measures show a significant rank-based correlation with contrast distance, when controlling for the other acuity measure. While the *r* values don't differ much from the linear correlations, this measure indicates that the relationship is not primarily driven by the outliers or greater variability. Furthermore, Fig. 5 shows that individuals with high auditory *and* somatosensory acuity produce the two fricatives with greater contrast between them than individuals with high acuity in only a single domain.

These results are consistent with prior studies of the sibilants (e.g., Perkell *et al.*, 2004b; Newman, 2003). In Perkell *et al.* (2004b) speakers' auditory acuity for synthetic sibilant ([s],[ʃ]) stimuli was found to be correlated with the degree of acoustic contrast they produced. Better sibilant contrast distances were also found in speakers who showed more consistent use of contact between the tongue tip and the lower alveolar ridge (a saturation effect) in producing [s] but not [ʃ]. Newman (2003) found significant correlations between the frequencies of spectral peaks of fricatives in participants' productions and those of the auditory stimuli they perceived to be prototypical of fricatives. The combined results of the previous and current studies provide converging evidence that speakers with higher auditory *and* somatosensory acuity will produce their sibilants with greater contrast. Although the correlations are significant and the hierarchical structure shown in Fig. 5 suggests a strong relationship between acuity and contrast, we cannot infer causality from such results. That is, discrimination may be one component of a possible set of factors that influences spoken contrast. However, these results are also consistent with the theoretical framework underlying the DIVA model of speech production (Guenther *et al.*, 2006). An interpretation of the results within the context of this mechanistic framework would suggest that high discrimination ability is a necessary, but perhaps not sufficient, requirement for producing clear sibilant contrasts.

In the DIVA framework, it is hypothesized that all speech sounds have associated somatosensory and auditory goals that are learned during speech acquisition. Infants acquire auditory goals relatively early in speech acquisition. The corresponding somatosensory goals are formed somewhat later, when the infant learns the somatosensory patterns that accompany his/her own successful productions of sounds. These somatosensory targets should be useful to the system especially when auditory feedback is disrupted either by the presence of environmental noise or hearing loss. If profound hearing loss occurs after robust fluent speech has been acquired, the use of somatosensory goals, which are not affected directly by hearing loss, may provide an important means of maintaining intelligibility. Most vowels and vowellike sounds are hypothesized to have more prominent auditory goals due to the relatively low amount of contact between tongue and palate and surrounding vocal-tract

structures, which would result in relatively limited somatosensory feedback. Therefore, vowel contrasts are hypothesized to rely more on auditory acuity. According to this line of thinking, consonants, which require articulatory contact, will obviously have more prominent somatosensory goals. Thus consonant production is hypothesized to be more strongly influenced by somatosensory acuity. The sibilants have prominent goals in both domains: turbulent noise with distinct acoustic spectra and patterns of articulatory position and contact. Sibilant contrasts should thereby rely on both somatosensory and auditory acuity. In the current study, the combination of both somatosensory and auditory acuity correlates significantly with the produced contrast distance, thus supporting the hypothesis of multisensory goals for sibilants (also see Perkell, 2010).

In a recent study, Shiller *et al.* (2009) used an apparatus that shifted the entire frequency spectrum downward during speakers' repeated, prolonged productions of [s]-initial utterances (e.g., sue). In response to the downward-shifted frequency of their /s/ productions, the subjects compensated by producing /s/ with an upward-shifted spectrum. The subjects' compensations persisted for a short time after the feedback perturbation was removed. These results provide additional evidence that one of the goals for sibilant production is in the auditory domain.

The findings of the current study and others (e.g., Villacorta *et al.*, 2007) lead to the inference that perceptual acuity has an important influence on the production of speech sounds and helps to account for some of the observed variability in the produced speech contrasts across speakers. Several further hypotheses can be drawn from the current findings and the theoretical framework of the DIVA model. Speakers with greater auditory discrimination will be more selective in their acceptance of perceived speech sounds as belonging to a particular phonemic category; therefore, speakers with higher auditory acuity will acquire auditory goal regions that are smaller and more precise than speakers with lower auditory acuity. Similarly, smaller, more precise somatosensory goal regions should be formed in speakers with high somatosensory acuity. This view leads to the general hypothesis that a child with higher acuity will not only acquire speech sounds with smaller goals regions, but will also produce these speech sounds with greater contrasts.

The results of this study should be interpreted within the limitations of the experiments to determine acuity. In the auditory discrimination experiment, a specific set of parameters derived from an individual speaker were used to synthesize the sibilant continuum, the same continuum was used across subjects and the auditory JND was evaluated at the category boundary. A groove-orientation identification paradigm was used to assess somatosensory acuity and the pressure on the probe and the timing of stimulus presentation, while maintained within a range, was manually controlled.

In summary, the results of the current study support the hypothesis that the contrast between the sibilants ([s] and [ʃ]) in a speaker's productions is positively correlated with both the auditory and the somatosensory discrimination capabilities of the speaker. These results also support the idea that sibilants have somatosensory and auditory targets and are consistent with results from prior studies (e.g., Newman, 2003; Perkell *et al.*, 2004a, 2004b). Finally, these results support the theoretical framework guiding the DIVA neurocomputational model of speech production (Guenther *et al.*, 2006), in which sensory feedback during speech production plays a key role in the acquisition and maintenance of motor programs for different speech sounds.

## ACKNOWLEDGMENTS

The study was supported by NIH Grant R01 DC01925 (Joseph Perkell, PI). We would like to thank Prof. Barbara Shinn-Cunningham for suggestions regarding the auditory acuity experiment.

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (**2008**). "A system for online dynamic perturbation of formant frequencies and results from perturbation of the mandarin triphthong /iau/," in Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France, pp. 65–68.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Gracco, V. L., and Abbs, J. H. (**1989**). "Sensorimotor characteristics of speech motor sequences," Exp. Brain Res. **75**, 586–598.

Guenther, F. H. (**1994**). "A neural network model of speech acquisition and motor equivalent speech production," Biol. Cybern. **72**, 43–53.

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., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

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

Hughes, O. M., and Abbs, J. H. (**1976**). "Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence," Phonetica **33**, 199–221.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of *F*0 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.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**, 227–229.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007**). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech Lang. Hear. Res. **50**, 913–927.

Maeyama, T., and Plattig, K. H. (**1989**). "Minimal two-point discrimination in human tongue and palate," Am. J. Otolaryngol. **10**, 342–344.

Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Burton, E., Vick, J., Lane, H., Tiede, M., and Zandipour, M. (**2008**). "Perception and production of /r/ allophones improve with hearing from a cochlear implant," J. Acoust. Soc. Am. **124**, 3191–3202.

Matthies, M. L., Svirsky, M. A., Lane, H. L., and Perkell, J. S. (**1994**). "A preliminary study of the effects of cochlear implants on the production of sibilants," J. Acoust. Soc. Am. **96**, 1367–1373.

McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). "Speech compensation to structural modifications of the oral cavity," J. Acoust. Soc. Am. **100**, 1093–1104.

McNutt, J. C. (**1975**). "Asymmetry in two-point discrimination on the

tongues of adults and children," J. Commun. Disord. **8**, 213–220.
McNutt, J. C. (**1977**). "Oral sensory and motor behaviors of children with /s/ or /r/ misarticulations," J. Speech Hear. Res. **20**, 694–703.
Nasir, S. M., and Ostry, D. J. (**2008**). "Speech motor learning in profoundly deaf adults," Nat. Neurosci. **11**, 1217–1222.
Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.
Niemi, M., Laaksonen, J.-P., Ojala, S., Aaltonen, O., and Happonen, R.-P. (**2006**). "Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/," Int. J. Oral Maxillofac Surg. **35**, 920–923.
Perkell, J. (**2009**). "Movement goals and feedforward and feedback mechanisms in speech production," in Proceedings of the Third International Symposium on Biomechanics, Human Function and Information Science, JAIST, Kanazawa, Japan.
Perkell, J., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J. C., and Zandipour, M. (**2001**). "Planning and auditory feedback in speech production," in *Speech Motor Control in Normal and Disordered Speech*, 4th International Speech Motor Conference, Nijmegen, The Netherlands, pp. 5–11.
Perkell, J. S. (**2010**). "Movement goals and feedback and feedforward control mechanisms in speech production," J. Neurolinguist. In press.
Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., and Burton, E. (**2007**). "Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users," J. Acoust. Soc. Am. **121**, 505–518.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233–272.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004a**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.
Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (**1993**). "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.
Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (**2004b**). "The distinctness of speakers' /s/–/s/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res. **47**, 1259–1269.
Purcell, D. W., and Munhall, K. G. (**2006**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Ringel, R. L., and Ewanowski, S. J. (**1965**). "Oral perception. I. Two-point discrimination," J. Speech Hear. Res. **8**, 389–398.
Sato, T., Okada, Y., Miyamoto, T., and Fujiyama, R. (**1999**). "Distributions of sensory spots in the hand and two-point discrimination thresholds in the hand, face and mouth in dental students," J. Physiol. Paris **93**, 245–250.
See supplementary material at http://dx.doi.org/10.1121/1.3493430 Document No. E-JASMAN-128-030011 for Klatt parameter files for the endpoints of the said-shed continuum. For more information, see www.aip.org/pubservs/epaps.html.
Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (**2009**). "Perceptual recalibration of speech sounds following speech motor learning," J. Acoust. Soc. Am. **125**, 1103–1113.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA), pp. 380–384.
Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–1443.
Van Boven, R. W., and Johnson, K. O. (**1994**). "The limit of tactile spatial resolution in humans: Grating orientation discrimination at the lip, tongue, and finger," Neurology **44**, 2361–2366.
Vick, J. C., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., and Zandipour, M. (**2001**). "Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**, 1257–1267.
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.
Wohlert, A. B. (**1996**). "Tactile perception of spatial stimuli on the lip surface by young and older adults," J. Speech Hear. Res. **39**, 1191–1198.
Wohlert, A. B., and Smith, A. (**1998**). "Spatiotemporal stability of lip movements in older adult speakers," J. Speech Lang. Hear. Res. **41**, 41–50.