



**HAL**  
open science

## Visual feedback of the tongue influences speech adaptation to a physical modification of the oral cavity

Guillaume Barbier, Ryme Merzouki, Mathilde Bal, Shari R Baum, Douglas M Shiller

### ► To cite this version:

Guillaume Barbier, Ryme Merzouki, Mathilde Bal, Shari R Baum, Douglas M Shiller. Visual feedback of the tongue influences speech adaptation to a physical modification of the oral cavity. *Journal of the Acoustical Society of America*, 2021, 150, pp.718 - 733. 10.1121/10.0005520 . hal-03919666

**HAL Id: hal-03919666**

**<https://hal.science/hal-03919666v1>**

Submitted on 3 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Visual feedback of the tongue influences speech adaptation to a physical modification of the oral cavity

Guillaume Barbier,<sup>1</sup> Ryme Merzouki,<sup>1</sup> Mathilde Bal,<sup>1</sup> Shari R. Baum,<sup>2</sup> and Douglas M. Shiller<sup>1,a)</sup>

<sup>1</sup>*École d'Orthophonie et d'Audiologie, Université de Montréal, Case Postale 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada*

<sup>2</sup>*School of Communication Sciences and Disorders, McGill University, 2001 McGill College Avenue, Suite 800, Montréal, Québec H3A 1G1, Canada*

### ABSTRACT:

Studies examining sensorimotor adaptation of speech to changing sensory conditions have demonstrated a central role for both auditory and somatosensory feedback in speech motor learning. The potential influence of *visual* feedback of oral articulators, which is not typically available during speech production but may nonetheless enhance oral motor control, remains poorly understood. The present study explores the influence of ultrasound visual feedback of the tongue on adaptation of speech production (focusing on the sound /s/) to a physical perturbation of the oral articulators (prosthesis altering the shape of the hard palate). Two visual feedback groups were tested that differed in the two-dimensional plane being imaged (coronal or sagittal) during practice producing /s/ words, along with a no-visual-feedback control group. Participants in the coronal condition were found to adapt their speech production across a broader range of acoustic spectral moments and syllable contexts than the no-feedback controls. In contrast, the sagittal group showed *reduced* adaptation compared to no-feedback controls. The results indicate that real-time visual feedback of the tongue is spontaneously integrated during speech motor adaptation, with effects that can enhance or interfere with oral motor learning depending on compatibility of the visual articulatory information with requirements of the speaking task. © 2021 Acoustical Society of America. <https://doi.org/10.1121/10.0005520>

(Received 4 February 2021; revised 6 June 2021; accepted 15 June 2021; published online 3 August 2021)

[Editor: Susanne Fuchs]

Pages: 718–733

### I. INTRODUCTION

Studies examining the *adaptation* of speech production to changing physical or sensory conditions (i.e., changes in motor control that act to reduce their negative impact) have demonstrated a significant capacity for sensorimotor plasticity in the control of oral movements as well as a central role for both auditory and somatosensory feedback in speech learning and development (e.g., Baum and McFarland, 1997; Houde and Jordan, 1998; Tremblay *et al.*, 2003). The potential influence of *visual* feedback of oral articulators, such as the tongue, which is not typically available during speech production but has been used as a tool to enhance the training of novel speech motor patterns, remains much less clear.

Historically, the role of visual input in speech has primarily been investigated in the context of the perception of *other* speakers. When observing a person speaking, visual information associated with movements of the face and mouth is readily integrated with the acoustic speech signal to influence speech perception. For example, vision of the face can improve the ability of listeners to decode a noisy or otherwise atypical (e.g., foreign accented) acoustic speech signal (Erber, 1975; Sumbly and Pollack, 1954) and can also enhance the perception of clearly audible speech signals (Arnold and Hill,

2001). Further, when auditory and visual-facial signals representing the production of different speech sounds are presented simultaneously to the listener (i.e., incongruent audiovisual speech stimuli), strong perceptual interactions can be observed that support a key role for the visual signal in speech perception (McGurk and MacDonald, 1976).

While visual input has been established to play an important role in speech perception, its possible role in the sensorimotor processes governing the *production* of speech sounds is less well understood. A role for vision in speech motor development is evidenced by studies of early blind individuals, who show differences from sighted individuals in the control of oral speech movements under a variety of speaking conditions, including simple vowel production (Ménard *et al.*, 2009; Turgeon *et al.*, 2020) and fast or clear speech (Ménard *et al.*, 2016a; Ménard *et al.*, 2016b), and in response to sensory perturbations impacting speech production (Ménard *et al.*, 2016c; Trudeau-Fisette *et al.*, 2017).

Perhaps the largest body of evidence indicating that speakers are able to integrate visual information in the processes of speech production comes from studies testing the practical applications of real-time visual feedback of the oral articulators (mainly the tongue) during speech training, focusing on the production of novel speech sounds in a second language (L2) and on the treatment of speech production disorders. Real-time visual representations of the tongue in the oral cavity are possible through a number of

<sup>a)</sup>Electronic mail: douglas.shiller@umontreal.ca

technologies, including electropalatography (EPG; registering the pattern of contact between the tongue and the hard palate), electromagnetic articulography (EMA, measuring the position of small wired sensors attached to the surface of the tongue), and ultrasound imaging [providing a continuous two-dimensional (2D) representation of the tongue surface]. In general terms, visual feedback-assisted speech training involves presenting the speaker with a real-time representation of the tongue on a monitor, along with a clearly identified visual articulatory goal associated with a particular speech sound. The visual goal varies in form depending on the imaging technology: a tongue-palate contact pattern for EPG, a 2D or three-dimensional (3D) sensor position for EMA, or a specific tongue shape or location in the oral cavity for ultrasound.

Studies suggest that the use of such visual feedback-based procedures can improve outcomes in the training of L2 sound production (Bliss *et al.*, 2018). For example, using an approach based on EMA, Suemitsu *et al.* (2015) demonstrated improvements in the production of an English vowel by native Japanese speakers, even in the absence of auditory feedback. Katz and Mehta (2015) also used an EMA-based visual feedback approach to train a novel non-native consonant in English speakers. Following a brief (25–30 min) practice period, improvements were noted in both kinematics (improved accuracy of articulation) and acoustics. L2 production training using ultrasound-based visual feedback has also shown positive results. Gick *et al.* (2008) observed improvement in the production of a challenging sound contrast (/l–l/) for a small group of Japanese speakers following only 30 min of practice using real-time ultrasound imaging of the tongue.

Considerable interest has also emerged in the use of real-time visual feedback of the tongue in the clinical treatment of persistent (i.e., treatment-resistant) speech disorders. Some of the earliest studies employed EPG, whereby the speaker is fitted with an acrylic dental appliance that covers the hard palate with an array of contact-sensitive electrodes to provide real-time visual feedback of tongue-palate contact patterns during speech production. Numerous studies have examined the application of EPG in the remediation of speech sound disorders in children and adults, including those associated with cleft-lip and palate (Gibbon and Hardcastle, 1989; Lee *et al.*, 2009; Michi *et al.*, 1986; Whitehill *et al.*, 1996), functional phonological and articulation disorders (Carter and Edwards, 2004; Dagenais *et al.*, 1994; Gibbon and Hardcastle, 1987; Hitchcock *et al.*, 2017; McAuliffe and Cornwell, 2008), neurological disorders (Gibbon *et al.*, 2003; Gibbon and Wood, 2003; Hardcastle *et al.*, 1987), and hearing impairment (Bacsfalvi *et al.*, 2007). While the majority of these studies involved a small number of participants, they nonetheless demonstrate across a wide range of clinical populations that real-time visual feedback of the tongue can be of potential benefit in speech production training.

In recent years, the use of visual articulatory feedback for the treatment of speech disorders has seen a major shift

toward the use of ultrasound imaging. Compared with EPG and EMA, ultrasound offers considerable advantages in terms of cost, versatility, and non-invasiveness (both EPG and EMA perturb speech movements and therefore require a period of acclimatization to achieve normal tongue motor patterns; see, e.g., McLeod and Searl, 2006), while also providing a more complete image of the tongue surface. A planar (2D) image of the tongue surface is obtained when the transducer is placed under the chin in either a mid-sagittal orientation (providing an image of the tongue surface along the midline) or in a coronal (or frontal) orientation (providing an image of the tongue surface laterally; see Fig. 4).

The addition of tasks focusing on tongue shape or position using real-time ultrasound during speech therapy has been shown to yield improved speech outcomes in children and adults with a variety of speech disorders, including those associated with developmental speech sound disorder (Adler-Bock *et al.*, 2007; Bressmann *et al.*, 2016; McAllister Byun *et al.*, 2014; Cleland *et al.*, 2015; Hitchcock and Byun, 2015; Preston *et al.*, 2019; see also Sugden *et al.*, 2019, for review,) childhood apraxia of speech (Preston *et al.*, 2013), cleft-lip and palate (Roxburgh *et al.*, 2016), and hearing impairment (Bacsfalvi, 2010; Bernhardt *et al.*, 2003). It should be noted, however, that considerable variability in outcomes among clinical patients has often been observed (e.g., Bernhardt *et al.*, 2008; Cleland *et al.*, 2019; Preston *et al.*, 2016; Sjolje *et al.*, 2016). Such variability may reflect inherent limitations in the technology. For example, ultrasound images of the tongue can be difficult to interpret because it can be unclear where exactly along the tongue contour the images are collected, given the potentially limited field-of-view and lack of clear anatomical landmarks (Mozaffari *et al.*, 2018; Preston *et al.*, 2017). Tongue visibility is less of a problem for avatar based EMA systems (Katz *et al.*, 2020).

Such clinical studies broadly indicate that speakers are capable of utilizing visual feedback of the tongue for the purpose of speech learning. Because of the inherent complexity of clinical research protocols, however, significant limitations remain in our understanding of the sensorimotor processes underlying the integration of visual feedback with speech learning and control. Outcomes in clinical studies do not simply reflect the influence of visual information on a speech learning task, but rather result from a complex interaction between an underlying speech production deficit (which may be sensory, motor, and/or cognitive/linguistic in nature) and the specific visual feedback-based training protocol, which requires the patient to visually match a phoneme-dependent (and sometimes speaker-dependent) target tongue shape or position, such as curving the tongue tip upward or retracting the tongue body. In addition to the real-time visual feedback of the tongue during such protocols, *verbal* feedback of performance is also typically provided from the speech-language pathologist, which may also be critical to the success of the treatment.

Critically, such training protocols differ from the way in which speech motor learning is believed to occur under

typical conditions (outside of L2 learning and treatment for severe speech motor disorders requiring the acquisition of completely new speech motor representations), which is characterized by the absence of conscious strategies, explicitly defined sensory target-matching, and verbal cues from an external teacher. This is highlighted in experimental studies of *sensorimotor adaptation* that involve introducing a sensory feedback perturbation during the production of otherwise normal words or phrases and then observing spontaneous, practice-related changes in motor patterns that gradually offset the effect of the perturbation (e.g., Baum and McFarland, 1997; Houde and Jordan, 1998; Lametti *et al.*, 2018). Such adaptation is presumed to reflect an implicit process of sensorimotor plasticity associated with the updating of internal models (neural mappings) that predict the relationship between motor commands and their sensory consequences (see, e.g., Krakauer and Mazzoni, 2011). Demonstrating that visual articulatory feedback can similarly be used *spontaneously* in the sensorimotor adaptation of oral speech movements in the absence of an explicitly defined visual target-matching task would greatly strengthen the existing evidence that real-time visual feedback of one's own articulator movements can directly influence speech motor learning in neurotypical speakers.

The purpose of the present study is to explore the extent to which typical adult speakers will spontaneously integrate real-time visual feedback of the oral articulators with the processes of sensorimotor learning of speech production. We combine real-time ultrasound imaging of the tongue with an experimental manipulation—involving a precise, physical alteration of the hard palate—known to induce adaptation of the oral articulators during speech production. The goal is to determine whether the availability of visual feedback of the tongue will influence the adaptation of tongue movements to the perturbation during a brief, intense period of speech practice. Importantly, in the current protocol, participants are *not* provided—or instructed to visually match—any specific articulatory target. Further absent from the current protocol is any verbal feedback of performance from the experimenter in relation to the position or shape of the tongue. Rather, visual feedback is provided only as a supplement to existing somatosensory feedback that may be used to monitor the position of the tongue. This contrasts with the vast majority of clinical and L2 training protocols involving visual feedback, in which subjects are explicitly instructed to visually match a tongue shape or position while attempting to produce the target speech sound.

The speech motor learning task employed here involves adaptation to a rigid prosthesis worn in the mouth that alters the shape of the hard palate immediately behind the upper incisors (the *alveolar* ridge; see Fig. 1). This alteration of palatal shape has been shown to disrupt the ability to produce the sibilant fricative /s/, which under typical speaking conditions involves maintaining a precise constriction between the tongue and palate in the alveolar region combined with a grooved tongue shape that directs the airstream toward the incisors. Following the initial perturbation,

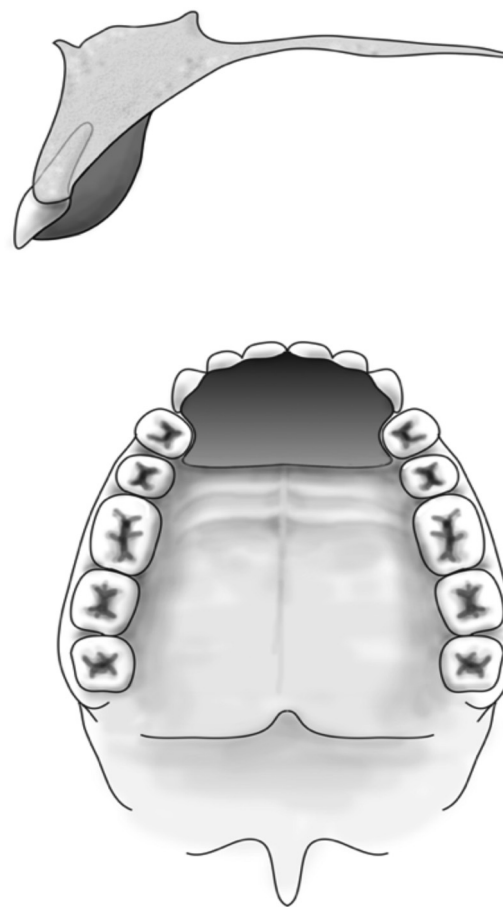


FIG. 1. Illustration of the placement of the palatal prosthesis (dark gray) in the mouth. Top image: Sagittal view of the upper palate and incisors. Bottom image: The palate and teeth, as viewed from below the maxillary region. Adapted from Baum and McFarland, *J. Acoust. Soc. Am.* 102, 2353–2359 (1997). Copyright 1997 AIP Publishing LLC (Baum and McFarland, 1997).

practice-related improvements in acoustic and articulatory patterns (i.e., sensorimotor adaptation) have consistently been observed (Aasland *et al.*, 2006; Barbier *et al.*, 2020; Baum and McFarland, 1997; Hamlet *et al.*, 1976; Thibeault *et al.*, 2011). Early studies of speech adaptation to a palatal prosthesis did not employ a strictly defined protocol of speech practice, but rather explored gradual improvements in speech output following an extended period of exposure (ranging from days to weeks; Hamlet *et al.*, 1976; Hamlet *et al.*, 1978). More recent studies have demonstrated significant improvements in speech acoustic properties as well as robust changes in tongue kinematic patterns following 15–20 min of focused speech practice with the prosthesis in place (Aasland *et al.*, 2006; Barbier *et al.*, 2020; Baum and McFarland, 1997; Thibeault *et al.*, 2011). Note that while /s/ has been the focus of the majority of studies involving speech adaptation to a palatal prosthesis (including the present one), a number of studies have shown that the perturbation also impacts the tongue movements associated with a range of consonant and vowel sounds (Barbier *et al.*, 2020; Brunner, 2009; Hamlet *et al.*, 1978; McFarland *et al.*, 1996).

In the present study, we examine the degree to which visual feedback of the tongue will influence the adaptation

of oral speech movements to a palatal prosthesis during production of the fricative /s/. We contrast the availability of two types of visual feedback of the tongue using 2D ultrasound with a third group that received no visual feedback during speech practice. The two visual feedback conditions differ with respect to the plane being imaged: *coronal* or *sagittal*. Both planes provide information about the tongue that is critical to the production of /s/. The coronal view provides a direct image of the central grooving of the tongue, an articulatory feature important for directing air toward the upper and lower central incisors. The sagittal view, in contrast, provides a direct image of tongue shape along the midline, as well as tongue position along the antero-posterior (front-back) and superior-inferior (up-down) axes, which determine the constriction location and the size and shape of the resonant cavity anterior to the constriction point—both key determinants of the fricative acoustic spectrum.

Changes in spectral properties of /s/ associated with the palatal perturbation and subsequent adaptation are measured in the current study by changes in the first four *spectral moments* (centroid, variance, skewness, and kurtosis), which together characterize the shape of the power spectral density of the signal. Spectral moments have long been recognized as stable acoustic correlates of fricative place of articulation (i.e., constriction location), in particular in distinguishing the sibilant fricatives /s/ and /ʃ/, with systematic differences in all four spectral moments revealed to varying degrees across a range of studies (Avery and Liss, 1996; Forrest *et al.*, 1988; Jongman *et al.*, 2000; McFarland *et al.*, 1996; Nissen and Fox, 2005; Nittrouer, 1995; Nittrouer *et al.*, 1989; Perkell *et al.*, 2004; Tjaden and Turner, 1997). As such, the spectral moments have served as the primary dependent measure in the majority of studies examining adaptation to palatal prostheses. While the majority of these studies have focused exclusively on the first spectral moment (Aasland *et al.*, 2006; Barbier *et al.*, 2020; Baum and McFarland, 1997, 2000), experimental effects involving higher moments have also been shown, including M2 (Thibeault *et al.*, 2011) and M3 and M4 (Brunner *et al.*, 2011; McFarland *et al.*, 1996). For completeness, all four spectral moments were examined in the present study.

Demonstrating that speakers show a benefit in sensorimotor adaptation outcomes when visual feedback of the tongue is made available in either (or both) of the ultrasound conditions would strengthen the limited existing evidence that real-time visual feedback of one's own articulator movements can influence speech motor learning in neurotypical speakers as well as more generally expand our understanding of how multiple sources of sensory feedback, including those that are not typically available during natural speech production, might be integrated during the learning and control of complex oral motor behaviors.

## II. METHODS

Forty-five native speakers of Quebec French (21–30 years of age) with no reported history of speech,

hearing, or language disorder were tested. To avoid large differences in vocal tract anatomy, all participants were female. The participants were all students in speech-language pathology at l'Université de Montréal and therefore had received some training in phonetics. Hearing status was assessed using pure-tone audiometry, verifying that the detection threshold in each ear was  $\leq 20$  dB hearing level (HL) at 0.5, 1, 2, 4, 6, and 8 kHz for all participants. Participants were randomly assigned to one of three ultrasound visual feedback conditions ( $n = 15$  in each group; see Sec. II C).

All procedures were approved by the Institutional Review Board of the Faculty of Medicine at l'Université de Montréal.

### A. Palatal prosthesis

The palatal prosthesis (Fig. 1) was custom fabricated for each participant using a biocompatible impression material (Express STD VPS, 3M, St. Paul, MN). An approximately 2 cm diameter ball of soft impression putty was gently pressed in place just behind the upper teeth until it self-hardened (1–2 min), at which point it was gently removed and hand-trimmed to meet the following dimensional specifications: 6 mm thickness behind the incisors, tapering off over a 1–2 cm distance ending at the first premolar (similar to the dimensions of palatal prostheses used in prior studies, e.g., Barbier *et al.*, 2020; Thibeault *et al.*, 2011). The prosthesis, which closely followed the contours of the alveolar region of the hard palate, was held in place using a thin layer of denture adhesive paste (Super Poligrip, GSK Consumer Healthcare, Brentford, UK) applied to the palatal surface.

### B. Speech production tasks

All speech tasks involved reading aloud a series of words or syllables presented one at a time on a 15-inch computer monitor located approximately 0.5 m in front of the participant. Participants carried out a series of four speech production tests in which they produced syllables containing the target consonant /s/ in combination with one of three possible vowels, /i/ (“ee”), /a/ (“ah”), and /u/ (“oo”), in two different syllable structures (consonant-vowel and vowel-consonant), yielding six different stimuli in total (/si, sa, su, is, as, us/). The three vowels were chosen for their association with tongue positions located near the limits of the French vowel production workspace (specifically, a high-front tongue position for /i/, a high-back position for /u/, and a low-central position for /a/). Each syllable was produced ten times in a randomized order, yielding 60 utterances per test.

The four speech production tests were carried out in the following sequence: (1) immediately preceding insertion of the prosthesis (test 1); (2) following insertion of the prosthesis, but before the period of speaking practice with the prosthesis in place (test 2); (3) following the speech practice



FIG. 2. (Color online) Schematic showing the sequence of insertion and removal of the palatal prosthesis (bottom), the series of speech tasks (middle), and the comparisons between speech tests that were the focus of the analyses in the current paper (top).

period with the prosthesis in place (test 3); and (4) following removal of the prosthesis (test 4; see Fig. 2).

With the palatal prosthesis in place, immediately following speech test 2, participants underwent a period of speech practice focusing on production of real French words whose initial sound was always the target /s/ followed by either the high-vowel /i/ (e.g., “Cigare”) or the low-vowel /a/ (e.g., “Sacré”). The six different syllable contexts in the speech production tests therefore permitted the examination of changes involving both practiced contexts (/si, sa/) and generalization to untrained vowel (/u/) and syllable contexts (vowel-consonant, /is, as, us/). During the practice period, a total of 15 different /si-/ words and 15 different /sa-/ words (see Table I) were presented two times each in a pseudo-randomized order (alternating between /si-/ words and /sa-/ words), for a total of 60 stimuli. Participants were instructed as follows: Following the visual presentation of each word on the monitor, which remained on screen for 3 s, participants had 20 s during which they were to produce the word 10 times. Their specific goal was to produce a typical-sounding /s/ at word onset, and participants were permitted to prolong their fricative production to achieve that. No instruction of any kind was given with regard to a desired shape or position of the tongue. Once 10 repetitions of the word were completed, participants were signaled visually to stop speaking until the 20-s practice window was complete,

TABLE I. French words (orthographic and phonemic transcriptions) used during the speech practice period, focusing on the word-initial /s/ sound in two vowel contexts.

/si-/ words		/sa-/ words	
Cigare	/sɪgar/	S’armer	/sarɛm/
Cime	/sim/	Sabot	/sabɔ/
Ciment	/simɑ̃/	Sabre	/sabr/
Circuit	/sɪrkɥi/	Sac	/sak/
Cire	/sir/	Sacré	/sakre/
Cirer	/sire/	Safran	/safirɑ̃/
Cirque	/sɪrk/	Sammy	/sami/
Civière	/sivjɛr/	Sapeur	/sapœr/
Civique	/sivik/	Sapin	/sapɛ̃/
Cyprés	/siprɛ/	Sarment	/sarmɑ̃/
Sibérie	/sibɛri/	Sarrau	/sarɔ/
Sien	/sjɛ̃/	Savant	/savɑ̃/
Simon	/simɔ̃n/	Saveur	/savœr/
Sirop	/siro/	Savoir	/savwar/
Syrie	/siri/	Savon	/savɔ̃/

at which point the next word appeared on screen. Following this protocol, participants produced a total of 600 /s/-initial words (300 /si-/ and 300 /sa-/) within a 20-min period. Stimulus presentation and data collection were controlled using custom software written in MATLAB (version 9.5; MathWorks, Natick, MA).

### C. Ultrasound visual feedback

Participants produced speech under three possible visual feedback conditions during the practice period: (1) no ultrasound visual feedback of the tongue (control group), (2) visual feedback of the tongue surface in the mid-sagittal plane (sagittal group), and (3) visual feedback of the tongue surface in the coronal (i.e., frontal) plane (coronal group).

Ultrasound imaging of the tongue surface was carried out on a personal computer (PC)-based ultrasound system (MicrUS EXT-1H, Telemed Medical Systems, Milan, Italy), using a 64-element convex transducer (20 mm radius, operating at 4 MHz) that was positioned under the participant’s chin. The system was controlled using the Echo Wave II software (Telemed Medical Systems), with B-mode imaging set to 80 mm depth and 92° field-of-view, yielding an image capture rate of ~80 Hz. Ultrasound gel (Aquasonic 100, BioMedical Instruments, Clinton, MI) was applied to the surface of the transducer prior to the orientation and practice periods and re-applied as needed throughout the experiment to maintain a consistent image of the tongue surface. Live (real-time) ultrasound images were presented on a 21-inch computer display (1920 × 1280 resolution, 60 Hz refresh rate) at a distance of 0.5 m, positioned just above the 15-inch display used to present the syllable/word stimuli for the speaking tasks (see Fig. 3).

The ultrasound transducer was rigidly attached to an adjustable microphone stand, which allowed the experimenter to adjust the position and angle of the transducer under the chin of the seated participant. For the *sagittal* view of the tongue [Fig. 4(A)], the transducer was visually

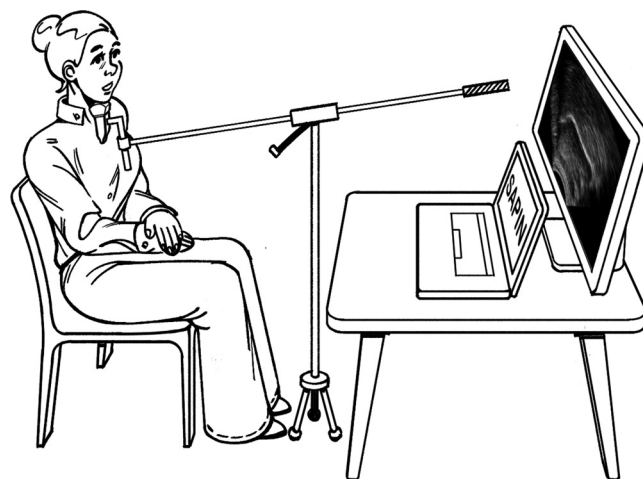
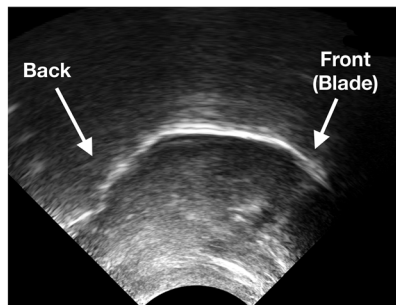


FIG. 3. Experimental setup. Illustration of the setup showing the relative position of the participant, ultrasound transducer, and computer displays for the visual word prompt and ultrasound visual feedback of the tongue.

(A) SAGITTAL View



(B) CORONAL View

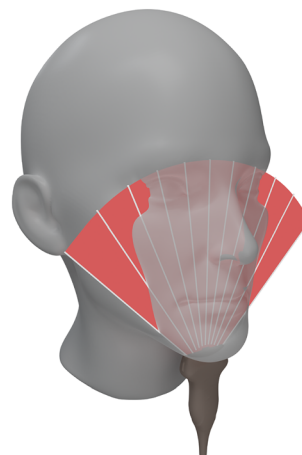
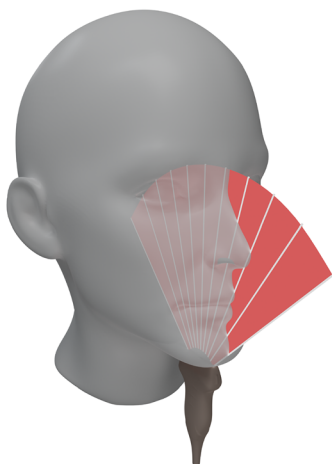
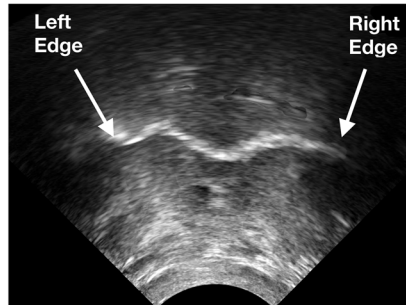


FIG. 4. (Color online) Ultrasound visual feedback. Examples of still images illustrating the real-time visual feedback of the tongue surface (appearing as a relatively bright curve extending from the left to right side of the screen) in the mid-sagittal (A) and coronal (B) planes, with a schematic below showing the orientation of the ultrasound transducer in each condition.

aligned with the participant’s midline under the chin and then slowly rotated forward and backward within the sagittal plane, such that the anterior portion of the tongue (i.e., the front/blade) was centered in the field-of-view when interacting with the alveolar region (identified by having the participant repeatedly produce the syllable “ta”). For the *coronal* view of the tongue [Fig. 4(B)], this exact same procedure was carried out, but then followed by a rotation of the transducer by 90° to align the transducer with the tongue front/blade in the coronal plane. The participant was then asked to produce a sustained /s/ sound, to verify that the entire (edge-to-edge) tongue surface was visible.

Following the above procedure for placing the transducer, the stand was locked firmly in position. While the stand helped to stabilize the position of the transducer, the participant was also permitted to hold the transducer gently with their hand to further reduce drift in the position over time as well as to allow the participant to adjust the level of pressure under the chin, if necessary. Transducer placement and ultrasound image quality were closely monitored by the experimenter throughout the experiment, and verbal instructions to the participant to make minor adjustments were provided as needed to maintain the best possible image quality.

Prior to the baseline speech test 1, all participants, including those in the control group, received a basic orientation regarding the ultrasound imaging system (transducer, gel, etc.) as well as how to interpret the images, including identification of the tongue surface (appearing as a bright line) and orientation of the image relative to the head (up/down/front/back for the sagittal group; up/down/left/right for the coronal group). Participants were then provided a

brief (~1-min) period of practice during which they produced several repetitions of the consonant-vowel sequences “ta” and “ka” to observe the effect on the image of the tongue under typical conditions (without the prosthesis in place).

During the 20-min practice period, participants in the two visual feedback groups were instructed to maintain visual fixation on the image of the tongue surface during their repeated production attempts. Importantly, no description or instruction of any kind was provided pertaining to the typical or expected tongue shape for the production of the target fricative /s/. Subjects in the control group also maintained the ultrasound transducer under their chin (in a sagittal orientation) for the duration of the 20-min practice period to match the physical sensation of the transducer under the chin.

Note that for participants in all three groups, the ultrasound transducer was positioned under the chin in a sagittal orientation during the four speech tests to record, for future study, the tongue movement patterns associated with the palatal perturbation and speech adaptation. The conditions of the recording were identical for all participants, and no ultrasound visual feedback was provided to any participants during these tests.

**D. Acoustic recording and analysis**

All signal recording and analysis was performed using custom routines written in MATLAB. The acoustic speech signal was digitized at 44.1 kHz (16-bit) using a cardioid microphone (C520, AKG, Hofgeismar, Germany) mounted

25 cm from the participant. For each recorded syllable produced in each of the four speech production tests, onset and offset of the fricative /s/ was identified on the basis of the RMS amplitude and then manually verified by visual inspection of the waveform. From the identified fricatives, a 40-ms window aligned at the fricative onset was used for subsequent spectral analysis. Focusing on the onset of sound production simplifies the interpretation of any observed changes in fricative acoustic properties associated with the palatal perturbation and subsequent speech practice, as it avoids the contribution of feedback-driven (online) corrective changes during the utterance (see, e.g., Niziolek *et al.*, 2013). Hence, any observed changes can be attributed to the learned (i.e., planned) control of the articulators associated with /s/ production.

For each 40-ms segment, the power spectral density was computed (*pmtm* function; Signal Processing Toolbox, version 8.4, MathWorks) using the Thompson multitaper method with eight tapers (Thomson, 1982). For random signals, such as frication noise, the multitaper method yields a lower variance estimate of the spectrum compared to the traditional discrete Fourier transform and has been used in a number recent studies involving the spectral analysis of fricatives (e.g., Koenig *et al.*, 2013; Todd *et al.*, 2011).

Changes in fricative spectra were examined by computing the first four moments of the spectral distribution (abbreviated M1–M4), which characterize the shape of the power spectral density of the signal. The four spectral moments correspond, respectively, to the frequency centroid (i.e., mean of the distribution; M1), variance (i.e., spread of the distribution; M2), skewness (i.e., asymmetry, which can be positive, indicating a longer right tail in the distribution, or negative, indicating a longer left tail; M3), and kurtosis (i.e., “tailedness” of the spectral distribution, where higher values correspond to more extreme values in both tails of the distribution; M4).

In total, each participant contributed 240 values for each of the four spectral moments (10 repetitions × 6 syllables × 4 speech tests). Outliers were removed from among the ten repetitions of each syllable produced during each speech test using the interquartile range rule (values exceeding the median ± 2 times the interquartile range). This procedure resulted in the removal of approximately 7% of data points overall.

### E. Statistical analyses

Statistical analyses focused on two separate experimental effects of interest: (1) the effect of *palate insertion* (difference between test 2 and test 1) and (2) the effect of *speech practice* with the palatal prosthesis in place (difference between test 3 and test 2), using a linear mixed-effects (LME) modeling approach in R (version 4.0.1; R Core Team, 2020) with *lme4* (version 1.1; Bates *et al.*, 2015). The focus on these two effects of interest is based directly upon our prior work (Barbier *et al.*, 2020) and represents a hypothesis-driven set of planned comparisons that does not

include the final speech test (test 4; following removal of the palate). The complete dataset, including all four speech tests, is destined for a planned future examination of tongue kinematics using the recorded ultrasound images from the current study.

For each of the two experimental effects of interest, the significance of changes in each of the four acoustic measures was tested by fitting the model

$$\text{Acoustic.Measure} \sim \text{GROUP} * \text{SYLLABLE} * \text{TEST} + (\text{TEST} | \text{Participant}), \quad (1)$$

where *Acoustic.Measure* corresponds to the spectral moment, *GROUP* corresponds to the three ultrasound visual feedback conditions (control, sagittal, and coronal, with control as the reference condition), *SYLLABLE* refers to the six stimuli (/si, sa, su, is, as, us/; with /si/ as the reference level), and *TEST* corresponds to the two speech production tests defining the experimental effect of interest (test 2 vs test 1 for the effect of insertion, and test 3 vs test 2 for the practice effect). Finally, (*TEST* | *Participant*) represents the inclusion of random intercepts per participant and of random slopes and intercepts of the effect of test per participant. Note that the model does not include random slopes for the effect of syllable, as their inclusion yields convergence errors. The significance of the fixed effects (including the two-way and three-way interactions) was evaluated using the R package *lmerTest* (version 3.1), which provides analysis of variance (ANOVA)-style significance tables using Satterthwaite’s degrees-of-freedom method. This allows for the reporting of readily interpretable degrees-of-freedom, *F*-values, and *p*-values. *Post hoc* comparisons between fixed effect levels were carried out when appropriate using *z*-tests on estimated marginal means using the R package *emmeans* (version 1.4.7) and applying the Holm–Bonferroni correction for multiple comparisons.

## III. RESULTS

### A. Baseline production

Baseline values of the four spectral moments associated with the production of /s/ in each context, averaged across all participants, are shown in Fig. 5. Overall, the values are in the range of those reported in previous studies (Jongman *et al.*, 2000; McFarland *et al.*, 1996; Nittrouer, 1995). While the acoustic properties of /s/ production among different vowel and syllable contexts are not the focus of the current study, the average values provide context in which to interpret the perturbing effect of the palatal prosthesis. As can be seen in Fig. 5, little systematic difference is observed between the two syllable types (consonant-vowel vs vowel-consonant); however, as reported previously (Jongman *et al.*, 2000), vowel context does show some influence. In particular, in the /u/ (high-back) vowel context, /s/ is characterized by lower average spectral mean, higher variance, less negative skewness, and lower kurtosis in comparison with /i/ and /a/ vowel contexts.



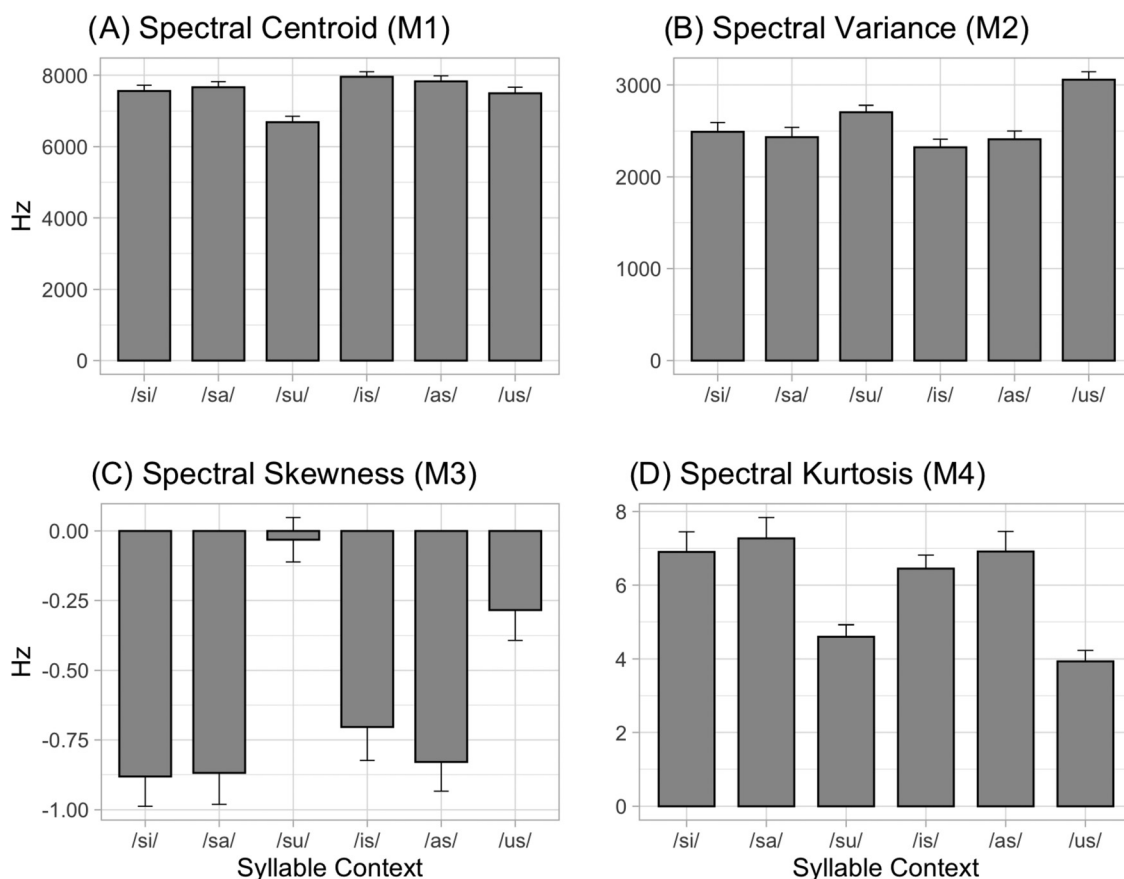


FIG. 5. Baseline production. Shown are average values of the four spectral moments associated with baseline /s/ production in the six different syllable contexts. Error bars,  $\pm 1$  standard error of the mean.

### B. Effect of insertion of the palatal prosthesis

Changes in the four spectral moments associated with the insertion of the palatal prosthesis, calculated as the difference between speech test 2 (immediately after insertion) and test 1 (baseline), are shown in Fig. 6. The impact of the prosthesis on /s/ production is characterized by a systematic decrease in spectral mean, an increase in spectral variance, more positive spectral skewness, and more negative spectral kurtosis. These patterns are consistent with those reported in prior studies involving /s/ production with a palatal prosthesis (e.g., Barbier *et al.*, 2020; McFarland *et al.*, 1996).

For each acoustic measure, a LME analysis was used to assess the effect of palate insertion (i.e., the fixed effect TEST) in combination with differences among the six syllable conditions (SYLLABLE) and the three visual feedback groups (GROUP). The results, including the LME model summary and an ANOVA-style table (using Satterthwaite’s method) reporting the significance of the main effects and the two- and three-way interactions as well as detailed results of *post hoc* comparisons (*z*- and *p*-values) are provided in the supplementary material.<sup>1</sup>

For the spectral centroid (M1), the main effects of TEST [F(1,42) = 98.5,  $p < 0.001$ ] and SYLLABLE [F(5,4933) = 69.80,  $p < 0.001$ ] were found to be significant, as well as the interactions between GROUP and SYLLABLE [F(10,4933) = 11.25,  $p < 0.001$ ] and between TEST and SYLLABLE

[F(5,4933) = 23.02,  $p < 0.001$ ] and the three-way interaction [F(10,4933) = 2.29,  $p < 0.05$ ]. *Post hoc* comparisons were carried out to assess the significance of the insertion effect (i.e., the effect of TEST) within each combination of syllable condition and visual feedback group. Results are summarized in Table II. The change in centroid following insertion of the palatal prosthesis was found to be statistically significant in all contexts for all groups ( $p < 0.05$ ).

For spectral variance (M2), the main effects of TEST [F(1,42) = 54.5,  $p < 0.001$ ], SYLLABLE [F(5,4912) = 88.9,  $p < 0.001$ ], and GROUP [F(2,42) = 8.43,  $p < 0.001$ ] were all significant, as were the interactions between GROUP and SYLLABLE [F(10,4912) = 8.1,  $p < 0.001$ ] and between SYLLABLE and TEST [F(5,4912) = 29.3,  $p < 0.001$ ] and the three-way interaction [F(10,4912) = 2.3,  $p < 0.05$ ]. *Post hoc* comparisons revealed a significant effect of TEST ( $p < 0.05$ ) in all but three cases: the syllable /su/ in the control group and the syllable /us/ in the coronal and sagittal groups (Table II).

For spectral skewness (M3), the main effects of TEST [F(1,42) = 39.0,  $p < 0.001$ ] and SYLLABLE [F(5,4889) = 125.9,  $p < 0.001$ ] were significant, as were the interactions between GROUP and SYLLABLE [F(10,4889) = 11.1,  $p < 0.001$ ] and between SYLLABLE and TEST [F(5,4889) = 19.49,  $p < 0.001$ ] and the three-way interaction [F(10,4889) = 2.0,  $p < 0.05$ ]. *Post hoc* comparisons

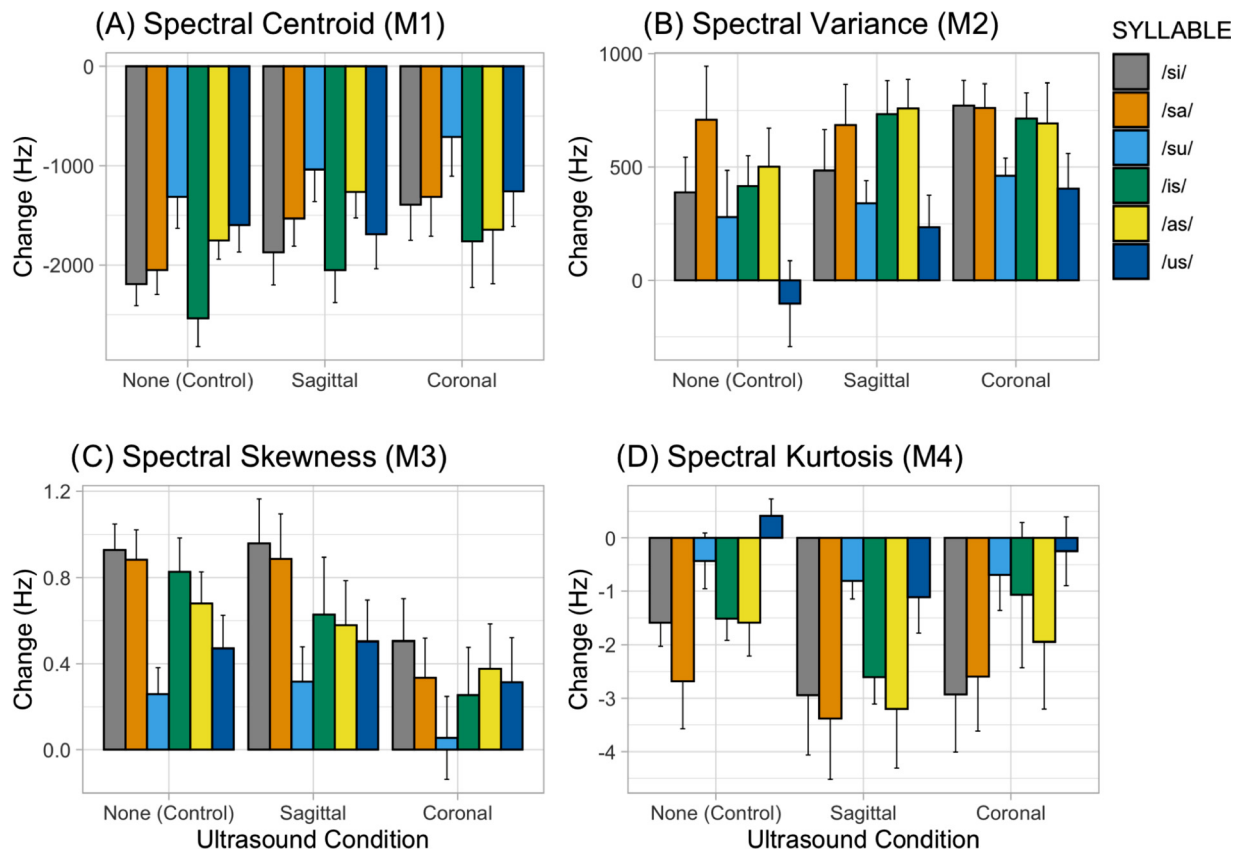


FIG. 6. (Color online) Insertion effect. Mean change in the four spectral moments associated with insertion of the palatal prosthesis is shown for each of the three visual feedback groups and each syllable context. Error bars,  $\pm 1$  standard error of the mean.

revealed a significant effect of TEST ( $p < 0.05$ ) in all but six cases: /su/ in all three groups and /sa/, /is/, and /us/ in the coronal group (Table II).

For spectral kurtosis (M4), the main effects of TEST [ $F(1,42) = 22.6, p < 0.001$ ], SYLLABLE [ $F(5,4792) = 68.2, p < 0.001$ ], and GROUP [ $F(2,42) = 4.4, p < 0.05$ ] were all significant, as well as the two-way interactions between GROUP and SYLLABLE [ $F(10,4792) = 4.7, p < 0.001$ ] and between SYLLABLE and TEST [ $F(5,4793) = 21.5, p < 0.001$ ]. *Post hoc* comparisons revealed a significant effect of TEST ( $p < 0.05$ ) in all but eight cases: /su/ and /us/ in all three groups, /is/ in the control and coronal groups, and /as/ in the control group (Table II).

TABLE II. Tests of insertion effect. Summary of *post hoc* pairwise evaluation of the difference between speech test 2 and test 1 (i.e., the insertion effect) in each syllable context for each of the three experimental groups. Rows show results for the four acoustic measures (M1–M4). Syllable contexts targeted in the practice phase (/si, sa/) are shown in bold. \*, a significant result ( $p < 0.05$ ). Detailed results are provided in the supplemental materials (see footnote 1).

GROUP: SYLLABLE:	Control		Sagittal			Coronal						
	si	sa	su	is	as	us	si	sa	su	is	as	us
M1:	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>
M2:	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>
M3:	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>
M4:	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>*</b>

In summary, insertion of the palatal prosthesis was associated with broad, systematic changes in all four spectral moments, as indicated by the significant main effect of TEST in each case. *Post hoc* tests revealed a reduced (non-significant) effect magnitude for certain syllable contexts, in particular, those involving the vowel /u/ (possibly due to the coarticulatory effect of a more retracted tongue posture or increased lip rounding from the vowel to the fricative), with some variation between the experimental groups. Importantly, however, the insertion effect was statistically significant and relatively large in magnitude for the syllable contexts targeted in the 20-min practice phase (/si-/ and /sa-/), as well as for these same vowels when produced in the syllable-final position (/is/ and /as/), across the three visual feedback groups.

### C. Effect of practice with the prosthesis in place

Changes in the four spectral moments associated with the 20-min period of practice with the prosthesis in place, calculated as the difference between speech test 3 (immediately after practice) and test 2 (immediately prior to practice), are shown in Figs. 7 and 8. The effect of practice is characterized by systematic changes that act to *reduce* the impact of the perturbation (i.e., adaptation) across the four spectral measures. This includes an *increase* in spectral mean (by 51%, 23%, and 69% of the perturbation magnitude on average for the control, sagittal, and coronal groups,

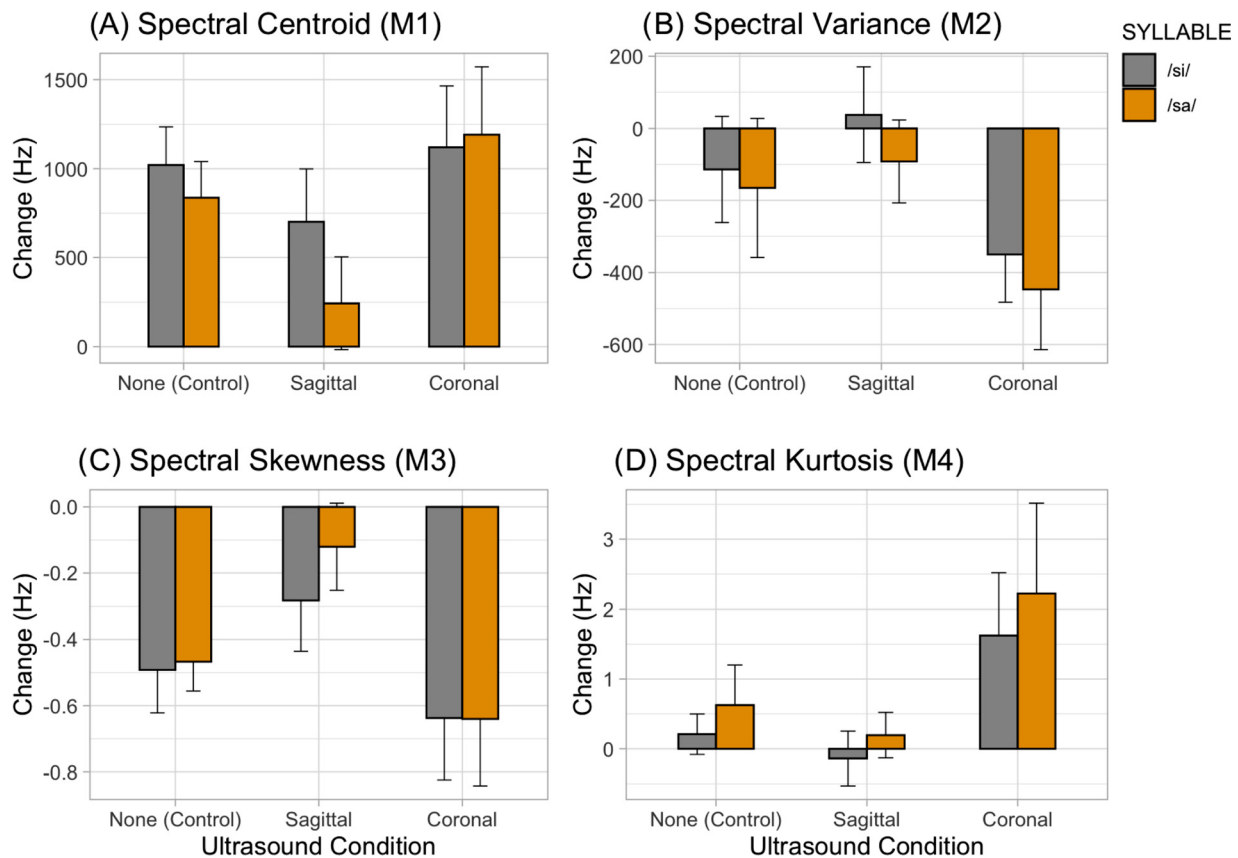


FIG. 7. (Color online) Practice effect in target syllable contexts. Mean change in the four spectral moments associated with the 20-min period of practice is shown for each of the three visual feedback groups and each of the trained syllable contexts (/si/ and /sa/). Error bars,  $\pm 1$  standard error of the mean.

respectively), a *decrease* in spectral variance (by 29%, 12%, and 52% for the three groups, respectively), more *negative* spectral skewness (by 68%, 22%, and 98%, respectively), and more *positive* spectral kurtosis (by 11%, 5%, and 49%, respectively).

A LME analysis was used to assess the effect of 20 min of speech practice with the prosthesis in place (the effect of TEST, comparing test 3 and test 2), in combination with differences among the syllable conditions (SYLLABLE) and among the three visual feedback groups (GROUP). Figure 7 shows the mean practice-related change (i.e., the TEST effect) for each of the four spectral moments in the two trained syllable contexts, while Fig. 8 shows the mean change in the four untrained contexts. Detailed results of the analyses, including model summaries and ANOVA-style tables, as well as detailed results of the *post hoc* tests, are provided in the supplementary material.<sup>1</sup>

For the spectral centroid, the main effects of TEST [F(1,42) = 34.5,  $p < 0.001$ ], SYLLABLE [F(5,4923) = 29.4,  $p < 0.001$ ], and GROUP [F(2,42) = 5.2,  $p < 0.01$ ] were all found to be significant, along with the two-way interaction between SYLLABLE and GROUP [F(10,4923) = 10.3,  $p < 0.001$ ]. The three-way interaction was also found to be marginal [F(10,4923) = 1.8,  $p = 0.052$ ]. To better comprehend the various main and interaction effects, while also addressing the key question of which conditions showed a significant effect of practice, *post hoc* comparisons were

carried out to assess the effect of TEST within each combination of syllable condition and visual feedback group. Results are summarized in Table III. For the targeted syllable contexts /si/ and /sa/, practice-related improvement was statistically significant ( $p < 0.05$ ) for the control group and the coronal group; however, for the sagittal group, only the change associated with production of /si/ was significant. Of the four untrained contexts, the control group showed a significant improvement in three contexts (/su, as, us;  $p < 0.05$ ), and the coronal group showed improvement in all four contexts ( $p < 0.05$ ), whereas the sagittal group failed to show a significant change in any context.

For spectral variance, the main effects of TEST [F(1,42) = 7.5,  $p < 0.01$ ], SYLLABLE [F(5,4923) = 24.8,  $p < 0.001$ ], and GROUP [F(2,42) = 8.2,  $p < 0.001$ ] were found to be significant, along with the two-way interaction between SYLLABLE and GROUP [F(10,4923) = 12.1,  $p < 0.001$ ]. The three-way interaction was also found to be marginal [F(10,4923) = 1.8,  $p = 0.051$ ]. *Post hoc* tests indicate that the change in the two targeted vowel contexts was significant only for the coronal group ( $p < 0.05$ ), with no significant results for either the control or sagittal groups. All remaining syllable contexts were non-significant for all three groups (Table III).

For spectral skewness, the main effects of TEST [F(1,41) = 38.42,  $p < 0.001$ ] and SYLLABLE [F(5,4886) = 97.3,  $p < 0.001$ ] were significant, along with the

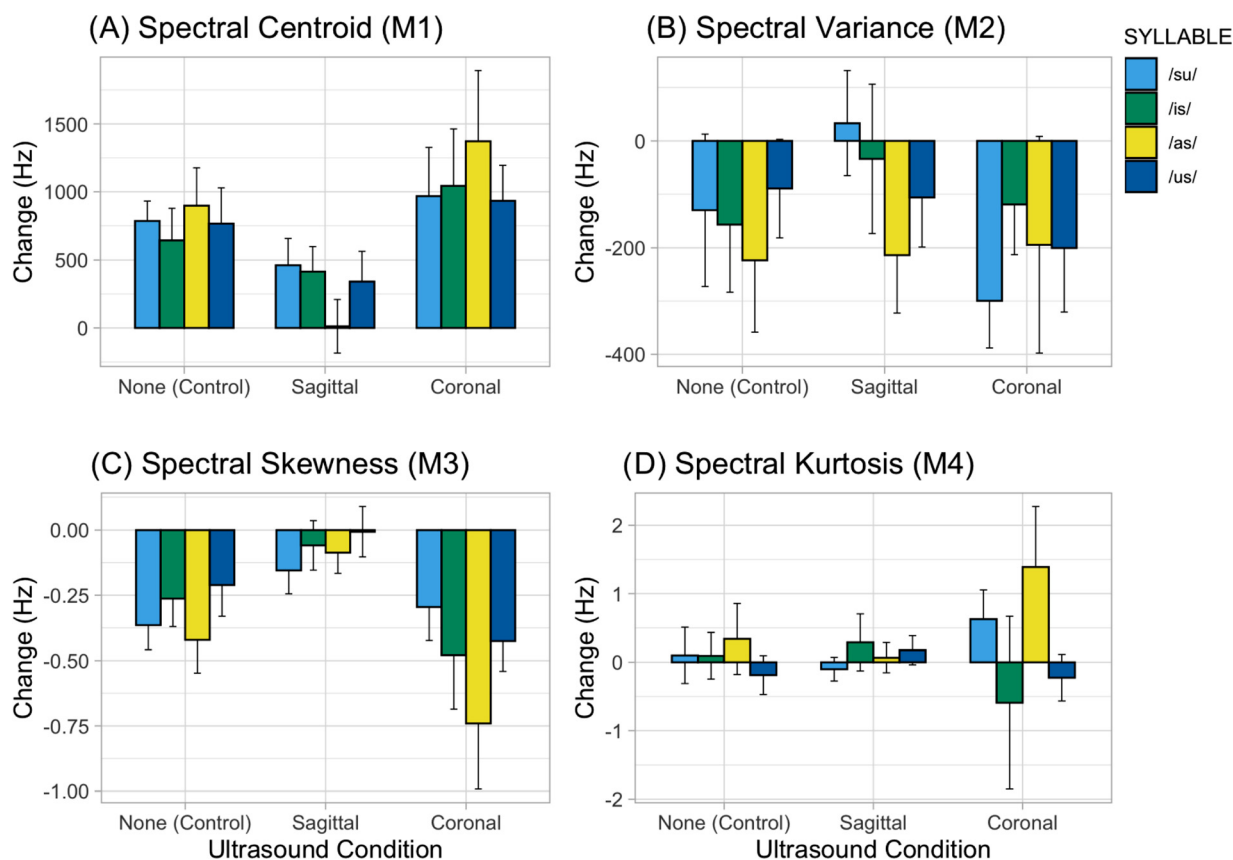


FIG. 8. (Color online) Practice effect in untrained syllable contexts. Shown is the mean practice-related effect in the four spectral moments in the four untrained syllable contexts /su/, /is/, /as/, and /us/. Error bars,  $\pm 1$  standard error of the mean.

interactions between TEST and GROUP [ $F(2,41) = 5.0, p < 0.05$ ], TEST and SYLLABLE [ $F(5,4887) = 5.9, p < 0.001$ ] and between GROUP and SYLLABLE [ $F(10,4886) = 7.7, p < 0.001$ ] and the three-way interaction [ $F(10,4887) = 2.0, p < 0.05$ ]. *Post hoc* tests indicate a significant change in the two targeted vowel contexts for both the control and coronal groups ( $p < 0.05$ ), but no significant improvement in either syllable context for the sagittal group. For the non-practiced syllable contexts, significant changes were shown for the control group in two contexts (/su, as/) and for the coronal group in three contexts (/is, as, us/), while the sagittal group showed no significant effects (Table III).

TABLE III. Tests of practice effect. Summary of *post hoc* pairwise evaluation of the difference between speech test 3 and test 2 (i.e., the practice effect) within each of the six syllable contexts for each of the three experimental groups. The rows show results for the four acoustic measures (M1–M4). The two syllable contexts targeted in the practice phase (/si/ and /sa/) are shown in bold. \*, a significant result ( $p < 0.05$ ). Detailed results are provided in the supplementary material (see footnote 1).

GROUP:	Control		Sagittal				Coronal					
	si	sa	su	is	as	us	si	sa	su	is	as	us
M1:	*	*	*	*	*	*	*	*	*	*	*	*
M2:									*	*		
M3:	*	*	*	*	*				*	*	*	*
M4:									*	*		

Finally, for spectral kurtosis, the main effects of SYLLABLE [ $F(5,4785) = 39.7, p < 0.001$ ] and GROUP [ $F(2,42) = 11.8, p < 0.001$ ] were significant, along with the interactions between TEST and SYLLABLE [ $F(5,4785) = 5.7, p < 0.001$ ] and GROUP and SYLLABLE [ $F(10,4785) = 14.7, p < 0.001$ ] and the three-way interaction [ $F(10,4785) = 3.7, p < 0.001$ ]. *Post hoc* tests showed a significant improvement in both targeted vowel contexts for the coronal group ( $p < 0.05$ ), but not for the control or sagittal groups. Changes in the four untrained contexts were all non-significant in all three groups (Table III).

Summarizing the changes observed in the two practiced syllable contexts (/si, sa/), participants in the no-visual-feedback control group demonstrated robust practice-related changes in spectral centroid (M1) and skewness (M3), but not in spectral variance (M2) or kurtosis (M4). In contrast, participants who received visual feedback of the tongue surface in the coronal plane exhibited a robust pattern of adaptation in /s/ production across all four spectral measures. Strikingly, participants who received visual feedback of the sagittal tongue surface showed a considerably more limited pattern of adaptation than both the coronal group and the control group, with a statistically significant effect noted only for the spectral centroid, and only in one context (/si/).

The four syllable contexts that were not targeted during the practice phase showed a more limited pattern of significant changes overall; however, differences between the

three groups in the pattern of compensation were still noted. The control group showed a statistically significant improvement in centroid frequency for three contexts and in skewness for two contexts. Similarly, the coronal group showed improvement in centroid for all four contexts and in skewness for three contexts. The sagittal group, however, showed no significant changes in any spectral measure for any of the unpracticed syllable contexts.

#### IV. DISCUSSION

In the present study, we examined whether the availability of ultrasound-based visual feedback of the tongue would influence speakers' spontaneous adaptation of oral movements to a palatal prosthesis affecting production of the fricative /s/. Two visual feedback groups were tested that differed with respect to the 2D plane being imaged (coronal and sagittal), along with a control group that received no visual feedback during speech training.

Insertion of the palatal prosthesis resulted in systematic changes across the four measured spectral moments: centroid, variance, skewness, and kurtosis. Following a 20-min period of speech practice focusing on words beginning with /si-/ and /sa-/, acoustic changes were assessed in the two trained contexts (/si, sa/), as well as four additional contexts to examine the generalization of training effects (/su, is, as, us/). For the two practiced contexts, participants in the coronal feedback group showed a robust pattern of adaptive changes opposing the effect of the perturbation on /s/ production across all four spectral measures. In contrast, the no-feedback control group showed improvements only in centroid and skewness. Strikingly, participants who received visual feedback of the tongue in the *sagittal* plane showed a more limited pattern of improvement than both the coronal and control groups, with significant improvement observed only in one acoustic measure (spectral centroid), and only in one context (/si/). The four syllable contexts that were not targeted during the practice phase showed a more restricted pattern of improvement overall; however, differences between the three groups were still noted. Changes in two measures—centroid and skewness—were observed in approximately half of the syllable contexts for both the coronal feedback group and the no-feedback control group. The sagittal group, however, showed no significant changes in any spectral measure for any of the unpracticed syllable contexts.

The finding that the coronal visual feedback group showed robust speech production improvements across a broader range of spectral measures and syllable contexts than the no-feedback control group supports the conclusion that ultrasound-based visual feedback of the tongue can enhance the sensorimotor adaptation of speech production, even in the absence of an explicitly defined visuospatial goal related to the speaking task.

A clear difference was also noted between the coronal and sagittal visual feedback conditions in the magnitude of the training effects. This was not a predicted result, as both

imaging axes provide information about the tongue that is known to be relevant to the production of the sibilant fricative /s/. Specifically, the coronal view shows the central grooving of the tongue, which is critical for channeling air toward the incisors, whereas the sagittal view shows the position and shape of the tongue along the midline, which determines the size and shape of the anterior resonating cavity (Ladefoged and Johnson, 2014; Shadle, 1990; Stone and Lundberg, 1996). The observed difference in outcomes between the two visual-feedback conditions, however, indicates that these two sources of visual information do *not*, in fact, contribute equally in the specific case of adapting tongue motor patterns to a palatal prosthesis.

It is possible that the more limited effect of the mid-sagittal view in the current study may have resulted, in part, from the reduced visibility of the tongue tip due to the shadow of the mandible. Note, however, that while the apex itself may have been hidden from view, careful placement of the transducer ensured that a large portion of the tongue surface remained visible, including the front/blade, which prior studies have shown to be significantly involved in adaptation to a palatal prosthesis (Barbier *et al.*, 2020; Thibeault *et al.*, 2011). Critically, the observed difference between the coronal and sagittal visual feedback conditions indicates that, rather than simply being a consequence of *any* (arbitrary) visual signal that correlates with the speech behavior, the specific, task-dependent information about the tongue provided by the visual image is serving a function in the motor learning process. This reduces the likelihood that general cognitive or attentional factors (e.g., associated with the shifting of the subject's attention toward an external visual representation of the tongue) are responsible for the effect of visual feedback on speech training (see, e.g., Freedman *et al.*, 2007).

While there are strong reasons to predict the potential utility of visual feedback related to both mid-sagittal and coronal tongue surface based on a general understanding of the articulatory nature of /s/ production (central groove, shape of the anterior cavity, etc.), the specific articulatory effect of the palatal perturbation and subsequent adaptation is more complex. In a recent study, Barbier *et al.* (2020) explored the effect of a similar palatal perturbation on tongue kinematic patterns across a range of speech sounds (including /s/) in nine adult speakers. Focusing on the mid-sagittal plane using electromagnetic articulography, the study indicated that insertion of the palate induced a significant change in sagittal tongue position and that following a period of practice, participants individually compensated by adjusting tongue position in a direction opposing that of the perturbation. However, the study also revealed that the precise *direction* of the articulatory change (across the three tongue sensors) was highly variable across the study participants, indicating that there was in fact no universal kinematic pattern of perturbation and compensation.

While the precise nature of the articulatory changes associated with the palatal perturbation and subsequent adaptation remains unclear (and, as described above, was

likely to have varied among speakers), the specific information contained within the visual representation of the tongue nonetheless appears to have been critical in determining whether the resulting impact on learning was facilitatory or detrimental. As the difference between the two feedback conditions pertained solely to representation of the tongue surface, it is reasonable to conclude that the visual feedback served as a source of information about the physical state of the speech motor system. In current models of speech motor control, knowledge of the current state of the system plays a key role in sensory feedback-based subsystems driving speech production and speech motor learning, including both an auditory and somatosensory pathway [see Parrell *et al.* (2019) for review]. Building upon the considerable body of evidence that visual input plays a major role in speech *perception*, the action model (ACT) of speech motor control includes a pathway for integrating visual input about a speaker's own articulator movements, in addition to auditory and somatosensory feedback (Kröger *et al.*, 2009, 2011; Katz and Mehta, 2015). A key characteristic of these model-based accounts of sensory-driven speech motor control, however, is the existence of a sensory *target* with which feedback is compared. In the present study, where no visual articulatory target was provided, it remains unclear by what mechanism the visual representation of the tongue influenced speech motor adaptation to the palatal perturbation. One possibility is that, without an explicit visual target, participants at the beginning of the practice period may simply have *not* made use of the visual feedback for the purpose of oral motor control, relying (as usual) on somatosensory and auditory feedback-based mechanisms. With exposure to the visual signal over a period of practice, however, participants may have independently established a visual sensory target via its association with somatosensory and auditory signals, at which point visual-based error-correcting mechanisms contributed to the process of speech adaptation. Such a process of visual target formation may also possibly be influenced, to some degree, by a speaker's prior knowledge of the articulatory basis of the target speech sound. A second possibility avoids altogether the requirement of a visual sensory target. Rather, it is possible that the restricted 2D view of the tongue surface in either the sagittal or coronal plane may have served to *constrain* the manner in which participants explored the articulatory workspace in their search for a tongue configuration that would improve the speech acoustic signal. In other words, during the practice phase, subjects in the sagittal and coronal groups may have tended to produce tongue kinematic patterns that were visible in their respective imaging planes (i.e., changes in elevation, protrusion, and curvature along the midline for the sagittal view and changes in elevation and lateral curvature for the coronal view). When this visual-feedback-driven articulatory constraint was aligned with the articulatory requirements of the speech adaptation task, the result was a more efficient process of speech adaptation. On the other hand, when the visual-based constraint was *not* aligned with the articulatory requirements of the task, the result was an

impaired process of speech adaptation. Future studies could directly test both of these possible scenarios (including the possibility that *both* may have played a role) by examining 3D kinematic measures of tongue motor patterns throughout the period of speech practice (e.g., using electromagnetic articulography), in combination with the different types of visual ultrasound feedback.

While the experimental protocol used in the current study differs in important ways from the application of visual feedback in the treatment of speech disorders, the current results nonetheless have implications for its use clinically. The finding of improved speech motor learning when ultrasound feedback was available (compared with the no-visual-feedback control group) broadly supports the use of this tool in the treatment of speech disorders in children and adults, for which the learning of new tongue motor patterns is often a principal goal (Duffy, 2019; Rvachew and Brosseau-Lapr e, 2016). However, the observation of a differential effect of ultrasound imaging plane (coronal vs sagittal) also raises a cautionary note about the manner in which visual feedback should be used in the training of different speech sounds. Specifically, the current results indicate that the information conveyed by the visual image of the tongue should be aligned with the articulatory requirements of the speech task. Notably, in the present study, the group receiving visual feedback in the sagittal plane showed speech adaptation that was *less* robust (across acoustic measures and syllable contexts) than the no-visual-feedback control group, suggesting that the presentation of feedback that is not optimized for the speech adaptation task may in fact have a detrimental effect on learning outcomes. Further study is clearly warranted to better understand the factors underlying the potentially negative impact of visual articulatory feedback on speech motor learning, for example, by examining the interaction between visual feedback and motor task across a much wider variety of feedback conditions and speech tasks.

The present study examined changes in the production of the fricative /s/ on the basis of an analysis of four spectral moments of the speech signal, a choice that was motivated by past work on the acoustics of fricative production. Producing a fricative requires maintaining a narrow constriction in the oral cavity, which creates airflow turbulence that acts as a source of broadband sound. The spectrum of this sound source is further shaped by interactions of the airstream with cavities and structures (e.g., the teeth or lips) anterior to the constriction (see, e.g., Stevens, 1998). The unvoiced alveolar fricative /s/, with its small anterior cavity and sibilant quality (i.e., airstream deflecting off of the teeth), is generally characterized by a well-defined (i.e., non-flat) spectrum with a frequency peak in the relatively high-frequency range (typically around 6 kHz for males and 7.5 kHz for females; Jongman *et al.*, 2000). No approach to characterizing the contrastive, perceptually salient acoustic features of /s/ has proven itself to be without limitations [see Koenig *et al.* (2013) for review]. However, spectral moments analysis has been shown to provide measures that

can reliably distinguish the alveolar /s/ from the alveopalatal sibilant /ʃ/ (i.e., the sound “sh”), as well as more broadly distinguish the sibilants (/s, ʃ/) from the non-sibilant fricatives /θ/ (“th”) and /f/. The spectral centroid has received the most attention, with numerous studies reporting a higher value for /s/ than for /ʃ/, likely owing to differences in the size of the anterior cavity (Jongman *et al.*, 2000; McFarland *et al.*, 1996; Nissen and Fox, 2005; Nittrouer *et al.*, 1989; Shadle and Mair, 1996; Tjaden and Turner, 1997). Systematic differences in spectral variance have been observed between the sibilant and non-sibilant fricatives (with sibilants showing lower values; Jongman *et al.*, 2000; Nissen and Fox, 2005; Shadle and Mair, 1996) and between /s/ and /ʃ/ (with /s/ showing a lower value; Tomiak, 1990). Spectral skewness has generally been shown to be more negative (i.e., tilted toward higher frequencies) for /s/ than other fricatives (Jongman *et al.*, 2000; McFarland *et al.*, 1996; Nissen and Fox, 2005; Nittrouer, 1995; Shadle and Mair, 1996). Finally, kurtosis has been shown to be higher (i.e., more peaked shape) for /s/ than other fricatives, including /ʃ/ (Jongman *et al.*, 2000; McFarland *et al.*, 1996), although some studies have shown a different pattern (Nissen and Fox, 2005). Interestingly, these differences between /s/ and /ʃ/ across the four spectral moments show some parallels with the perturbing effect of the palatal prosthesis on /s/ production. Specifically, compared to /s/, /ʃ/ is characterized by a lower spectral mean, greater variance, more positive skewness, and smaller kurtosis, matching the four effects of the prosthesis on /s/ observed in the present study. The articulatory basis of these spectral changes is likely different in these two situations, owing to the complex non-linear relationship between acoustics, tongue position, and palatal shape (see, e.g., Barbier *et al.*, 2020). Nonetheless, the similarities in acoustic effects further support the use of the four spectral moments to characterize the palatal perturbation and subsequent adaptation.

In summary, the present study explored the degree to which visual feedback of the tongue would be spontaneously used during sensorimotor adaptation of speech production to a physical oral perturbation. The results indicate that ultrasound-based visual feedback of the tongue can enhance the sensorimotor adaptation of speech production, even in the absence of an explicitly defined visual articulatory target or external verbal feedback about performance. However, it appears that such visual feedback may also *interfere* with sensorimotor adaptation, yielding weaker adaptation effects than a control condition involving no visual feedback, if the visual articulatory information is incompatible with the requirements of the speaking task.

## ACKNOWLEDGMENTS

This study was supported by grants from the Natural Sciences and Engineering Council of Canada (NSERC-Canada) and the Centre for Research on Brain, Language and Music (CRBLM). We thank Sabine Burfin for the illustrations in Fig. 3 and Noah Lebreque for the 3D

illustrations in Fig. 4. We also thank Ben Parrell and the anonymous reviewers for their valuable contributions.

<sup>1</sup>See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0005520> for complete LME model summaries and ANOVA-style tables, as well as detailed results of all post-hoc tests.

- Aasland, W. A., Baum, S. R., and McFarland, D. H. (2006). “Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation,” *J. Acoust. Soc. Am.* **119**, 2372–2381.
- Adler-Bock, M., Bernhardt, B. M., Gick, B., and Bacsfalvi, P. (2007). “The use of ultrasound in remediation of North American English /r/ in 2 adolescents,” *Am. J. Speech. Lang. Pathol.* **16**, 128–139.
- Arnold, P., and Hill, F. (2001). “Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact,” *Br. J. Psychol.* **92**, 339–355.
- Avery, J. D., and Liss, J. M. (1996). “Acoustic characteristics of less-masculine-sounding male speech,” *J. Acoust. Soc. Am.* **99**, 3738–3748.
- Bacsfalvi, P. (2010). “Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants,” *Can. J. Speech Lang. Pathol. Audiol.* **34**, 206–217.
- Bacsfalvi, P., Bernhardt, B. M., and Gick, B. (2007). “Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment,” *Adv. Speech Lang. Pathol.* **9**, 36–45.
- Barbier, G., Baum, S. R., Ménard, L., and Shiller, D. M. (2020). “Sensorimotor adaptation across the speech production workspace in response to a palatal perturbation,” *J. Acoust. Soc. Am.* **147**, 1163–1178.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). “Fitting linear mixed-effects models using lme4,” *J. Stat. Softw.* **67**, 1–48.
- 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.
- Bernhardt, B. M., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., O’connell, M., Sirianni, J., and Radanov, B. (2008). “Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada,” *Clin. Linguist. Phon.* **22**, 149–162.
- Bernhardt, B. M., Gick, B., Bacsfalvi, P., and Ashdown, J. (2003). “Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners,” *Clin. Linguist. Phon.* **17**, 199–216.
- Bliss, H., Abel, J., and Gick, B. (2018). “Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review,” *J. Second Lang. Pronunciation* **4**, 129–153.
- Bressmann, T., Harper, S., Zhylich, I., and Kulkarni, G. V. (2016). “Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors,” *Clin. Linguist. Phon.* **30**, 345–362.
- Brunner, J. (2009). “Perturbed speech: How compensation mechanisms can inform us about 1154 phonemic targets,” *Sudwestdeutscher Verlag Fur Hochschulschriften* **196**, hal-00372151; available at <https://hal.archives-ouvertes.fr/hal-00372151>.
- Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., and Perkell, J. (2011). “The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation,” *J. Speech Lang. Hear. Res.* **54**, 727–739.
- Carter, P., and Edwards, S. (2004). “EPG therapy for children with longstanding speech disorders: Predictions and outcomes,” *Clin. Linguist. Phon.* **18**, 359–372.
- Cleland, J., Scobbie, J. M., Roxburgh, Z., Heyde, C., and Wrench, A. (2019). “Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback,” *J. Speech Lang. Hear. Res.* **62**, 229–246.
- Cleland, J., Scobbie, J. M., and Wrench, A. A. (2015). “Using ultrasound visual biofeedback to treat persistent primary speech sound disorders,” *Clin. Linguist. Phon.* **29**, 575–597.
- Dagenais, P. A., Critz-Crosby, P., and Adams, J. B. (1994). “Defining and remediating persistent lateral lisps in children using electropalatography,” *Am. J. Speech. Lang. Pathol.* **3**, 67–76.

- Duffy, J. R. (2019). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 4th ed. (Elsevier Health Sciences, Philadelphia, PA).
- Erber, N. P. (1975). "Auditory-visual perception of speech," *J. Speech Hear. Disord.* **40**, 481–492.
- 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.
- Freedman, S. E., Maas, E., Caligiuri, M. P., Wulf, G., and Robin, D. A. (2007). "Internal versus external: Oral-motor performance as a function of attentional focus," *J. Speech Lang. Hear. Res.* **50**, 131–136.
- Gibbon, F., and Hardcastle, W. (1987). "Articulatory description and treatment of 'lateral/s/' using electropalatography: A case study," *Int. J. Lang. Commun. Disord.* **22**, 203–217.
- Gibbon, F., and Hardcastle, W. (1989). "Deviant articulation in a cleft palate child following late repair of the hard palate: A description and remediation procedure using electropalatography (EPG)," *Clin. Linguist. Phon.* **3**, 93–110.
- Gibbon, F., McNeill, A. M., Wood, S. E., and Watson, J. M. M. (2003). "Changes in linguopalatal contact patterns during therapy for velar fronting in a 10-year-old with Down's syndrome," *Int. J. Lang. Commun. Disord.* **38**, 47–64.
- Gibbon, F., and Wood, S. E. (2003). "Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: A case study," *Clin. Linguist. Phon.* **17**, 365–374.
- Gick, B., Bernhardt, B. M., Bacsfalvi, P., and Wilson, I. (2008). "Ultrasound imaging applications in second language acquisition," in *Phonology and Second Language Acquisition*, edited by J. Hansen and M. Zampini (John Benjamins, Amsterdam), Chap. 11, pp. 309–322.
- Hamlet, S., Geoffrey, V. C., and Bartlett, D. M. (1976). "Effect of a dental prosthesis on speaker-specific characteristics of voice," *J. Speech Hear. Res.* **19**, 639–650.
- Hamlet, S., Stone, M., and McCarty, T. (1978). "Conditioning prostheses viewed from the standpoint of speech adaptation," *J. Prosthet. Dent.* **40**, 60–66.
- Hardcastle, W. J., Morgan Barry, R. A., and Clark, C. J. (1987). "An instrumental phonetic study of lingual activity in articulation-disordered children," *J. Speech. Hear. Res.* **30**, 171–184.
- Hitchcock, E. R., and Byun, T. M. (2015). "Enhancing generalisation in biofeedback intervention using the challenge point framework: A case study," *Clin. Linguist. Phon.* **29**, 59–75.
- Hitchcock, E. R., Byun, T. M., Swartz, M., and Lazarus, R. (2017). "Efficacy of electropalatography for treating misarticulation of /t/," *Am. J. Speech. Lang. Pathol.* **26**, 1141–1158.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Katz, W., Lawn, A., and Kumar, H. (2020). "Opti-Speech: A real-time tongue model for research and speech training," *Proceedings of the 12th International Seminar on Speech Production (ISSP)*, December 14–18.
- Katz, W. F., and Mehta, S. (2015). "Visual feedback of tongue movement for novel speech sound learning," *Front. Hum. Neurosci.* **9**, 612.
- Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (2013). "Toward improved spectral measures of /s/: Results from adolescents," *J. Speech Lang. Hear. Res.* **56**, 1175–1189.
- Krakauer, J. W., and Mazzoni, P. (2011). "Human sensorimotor learning: Adaptation, skill, and beyond," *Curr. Opin. Neurobiol.* **21**, 636–644.
- Kröger, B. J., Kannampuzha, J., and Neuschaefer-Rube, C. (2009). "Towards a neurocomputational model of speech production and perception," *Speech Communication* **51**(9), 793–809.
- Kroger, B., Miller, N., Lowit, A., Lowit, and Neuschaefer-Rube, C. (2011). "Defective neural motor speech mappings as a source for apraxia of speech: Evidence from a quantitative neural model of speech processing (ACT)," in *Assessment of Motor Speech Disorders* (Plural Publishing), pp. 325–346.
- Ladefoged, P., and Johnson, K. (2014). *A Course in Phonetics* (Nelson Education, Toronto, Canada).
- Lametti, D. R., Smith, H. J., Watkins, K. E., and Shiller, D. M. (2018). "Robust sensorimotor learning during variable sentence-level speech," *Curr. Biol.* **28**, 3106–3113.e2.
- Lee, A. S.-Y., Law, J., and Gibbon, F. E. (2009). "Electropalatography for articulation disorders associated with cleft palate," *Cochrane Database Syst. Rev.* **3**, CD006854.
- McAllister Byun, T. M., Hitchcock, E. R., and Swartz, M. T. (2014). "Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention," *J. Speech Lang. Hear. Res.* **57**, 2116–2130.
- McAuliffe, M. J., and Cornwell, P. L. (2008). "Intervention for lateral /s/ using electropalatography (EPG) biofeedback and an intensive motor learning approach: A case report," *Int. J. Lang. Commun. Disord.* **43**, 219–229.
- 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.
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature* **264**, 746–748.
- McLeod, S., and Searl, J. (2006). "Adaptation to an electropalatograph palate: Acoustic, impressionistic, and perceptual data," *Am. J. Speech Lang. Pathol.* **15**, 192–206.
- Ménard, L., Côté, D., and Trudeau-Fisette, P. (2016a). "Maintaining distinctiveness at increased speaking rates: A comparison between congenitally blind and sighted speakers," *Folia Phoniatr. Logop.* **68**, 232–238.
- Ménard, L., Dupont, S., Baum, S. R., and Aubin, J. (2009). "Production and perception of French vowels by congenitally blind adults and sighted adults," *J. Acoust. Soc. Am.* **126**, 1406–1414.
- Ménard, L., Trudeau-Fisette, P., Côté, D., and Turgeon, C. (2016b). "Speaking clearly for the blind: Acoustic and articulatory correlates of speaking conditions in sighted and congenitally blind speakers," *PLoS One* **11**, e0160088.
- Ménard, L., Turgeon, C., Trudeau-Fisette, P., and Bellavance-Courtemanche, M. (2016c). "Effects of blindness on production-perception relationships: Compensation strategies for a lip-tube perturbation of the French [u]," *Clin. Linguist. Phon.* **30**, 227–248.
- Michi, K., Suzuki, N., Yamashita, Y., and Imai, S. (1986). "Visual training and correction of articulation disorders by use of dynamic palatography: Serial observation in a case of cleft palate," *J. Speech Hear. Disord.* **51**, 226–238.
- Mozaffari, M. H., Guan, S., Wen, S., Wang, N., and Lee, W. (2018). "Guided learning of pronunciation by visualizing tongue articulation in ultrasound image sequences," *Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, June 12–13, Ottawa, Canada.
- Nissen, S. L., and Fox, R. A. (2005). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," *J. Acoust. Soc. Am.* **118**, 2570–2578.
- Nittrouer, S. (1995). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *J. Acoust. Soc. Am.* **97**, 520–530.
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," *J. Speech Hear. Res.* **32**, 120–132.
- Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (2013). "What does motor efference copy represent? Evidence from speech production," *J. Neurosci.* **33**, 16110–16116.
- Parrell, B., Lammert, A. C., Ciccarelli, G., and Quatieri, T. F. (2019). "Current models of speech motor control: A control-theoretic overview of architectures and properties," *J. Acoust. Soc. Am.* **145**, 1456–1481.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004). "The distinctness of speakers' /s—/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," *J. Speech. Lang. Hear. Res.* **47**, 1259–1269.
- Preston, J. L., Brick, N., and Landi, N. (2013). "Ultrasound biofeedback treatment for persisting childhood apraxia of speech," *Am. J. Speech. Lang. Pathol.* **22**, 627–643.
- Preston, J. L., Maas, E., Whittle, J., Leece, M. C., and McCabe, P. (2016). "Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia," *Clin. Linguist. Phon.* **30**, 363–381.
- Preston, J. L., McAllister, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., and Whalen, D. H. (2017). "Ultrasound



- images of the tongue: A tutorial for assessment and remediation of speech sound errors," *J. Vis. Exp.* **119**, e55123.
- Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., and Whalen, D. H. (2019). "Remediating residual rhotic errors with traditional and ultrasound-enhanced treatment: A single-case experimental study," *Am. J. Speech. Lang. Pathol.* **28**, 1167–1183.
- R Core Team (2020). "R: A language and environment for statistical computing," version 4.0.1 (R Foundation for Statistical Computing, Vienna, Austria), <https://www.R-project.org/> (Last viewed September 1, 2020).
- Roxburgh, Z., Cleland, J., and Scobbie, J. M. (2016). "Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate," *Clin. Linguist. Phon.* **30**, 398–415.
- Rvachew, S., and Brosseau-Lapr e, F. (2016). *Developmental Phonological Disorders: Foundations of Clinical Practice*, 2nd ed. (Plural Publishing, San Diego, CA).
- Shadle, C. H. (1990). "Articulatory-Acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Springer, Dordrecht, Netherlands), pp. 187–209.
- Shadle, C. H., and Mair, S. J. (1996). "Quantifying spectral characteristics of fricatives," in *Proceedings of the Fourth International Conference on Spoken Language Processing: ICSLP '96*, October 3–6, Philadelphia, PA, Vol. 3, pp. 1521–1524.
- Sjolie, G. M., Leece, M. C., and Preston, J. L. (2016). "Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback," *J. Commun. Disord.* **64**, 62–77.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT, Cambridge, MA).
- Stone, M., and Lundberg, A. (1996). "Three-dimensional tongue surface shapes of English consonants and vowels," *J. Acoust. Soc. Am.* **99**, 3728–3737.
- Suemitsu, A., Dang, J., Ito, T., and Tiede, M. (2015). "A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning," *J. Acoust. Soc. Am.* **138**, EL382–EL387.
- Sugden, E., Lloyd, S., Lam, J., and Cleland, J. (2019). "Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders," *Int. J. Lang. Commun. Disord.* **54**, 705–728.
- Sumby, W. H., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.* **26**, 212–215.
- Thibeault, M., M enard, L., Baum, S. R., Richard, G., and McFarland, D. H. (2011). "Articulatory and acoustic adaptation to palatal perturbation," *J. Acoust. Soc. Am.* **129**, 2112–2120.
- Thomson, D. J. (1982). "Spectrum estimation and harmonic analysis," *Proc. IEEE* **70**(9), 1055–1096.
- Tjaden, K., and Turner, G. S. (1997). "Spectral properties of fricatives in amyotrophic lateral sclerosis," *J. Speech Lang. Hear. Res.* **40**, 1358–1372.
- Todd, A. E., Edwards, J. R., and Litovsky, R. Y. (2011). "Production of contrast between sibilant fricatives by children with cochlear implants," *J. Acoust. Soc. Am.* **130**, 3969–3979.
- Tomiak, G. R. (1990). "An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents," Ph.D. thesis, SUNY Buffalo, Buffalo, NY.
- Tremblay, S., Shiller, D. M., and Ostry, D. J. (2003). "Somatosensory basis of speech production," *Nature* **423**, 866–869.
- Trudeau-Fisette, P., Tiede, M., and M enard, L. (2017). "Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: Acoustic and articulatory data," *PLoS One* **12**, e0180300.
- Turgeon, C., Trudeau-Fisette, P., Lepore, F., Lipp e, S., and M enard, L. (2020). "Impact of visual and auditory deprivation on speech perception and production in adults," *Clin. Linguist. Phon.* **34**, 1061–1087.
- Whitehill, T. L., Stokes, S. F., and Yonnie, M. Y. (1996). "Electropalatography treatment in an adult with late repair of cleft palate," *Cleft Palate Craniofac. J.* **33**, 160–168.