

Research Article

The Relation of Articulatory and Vocal Auditory–Motor Control in Typical Speakers

Rosemary A. Lester-Smith,^a Ayoub Daliri,^a Nicole Enos,^{b,c} Defne Abur,^a Ashling A. Lupiani,^a Sophia Letcher,^d and Cara E. Stepp^{a,b,e}

Purpose: The purpose of this study was to explore the relationship between feedback and feedforward control of articulation and voice by measuring reflexive and adaptive responses to first formant (F_1) and fundamental frequency (f_o) perturbations. In addition, perception of F_1 and f_o perturbation was estimated using passive (listening) and active (speaking) just noticeable difference paradigms to assess the relation of auditory acuity to reflexive and adaptive responses.

Method: Twenty healthy women produced single words and sustained vowels while the F_1 or f_o of their auditory feedback was suddenly and unpredictably perturbed to assess reflexive responses or gradually and predictably perturbed to assess adaptive responses.

Results: Typical speakers' reflexive responses to sudden perturbation of F_1 were related to their adaptive responses

to gradual perturbation of F_1 . Specifically, speakers with larger reflexive responses to sudden perturbation of F_1 had larger adaptive responses to gradual perturbation of F_1 . Furthermore, their reflexive responses to sudden perturbation of F_1 were associated with their passive auditory acuity to F_1 such that speakers with better auditory acuity to F_1 produced larger reflexive responses to sudden perturbations of F_1 . Typical speakers' adaptive responses to gradual perturbation of F_1 were not associated with their auditory acuity to F_1 . Speakers' reflexive and adaptive responses to perturbation of f_o were not related, nor were their responses related to either measure of auditory acuity to f_o .

Conclusion: These findings indicate that there may be disparate feedback and feedforward control mechanisms for articulatory and vocal error correction based on auditory feedback.

^aDepartment of Speech, Language, & Hearing Sciences, Boston University, MA

^bDepartment of Biomedical Engineering, Boston University, MA

^cDepartment of Electrical & Computer Engineering, Boston University, MA

^dDepartment of Neuroscience, Kenyon College, Gambier, OH

^eDepartment of Otolaryngology—Head & Neck Surgery, Boston University School of Medicine, MA

Correspondence to Rosemary A. Lester-Smith, who is now at the Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin: rosemary.lester-smith@austin.utexas.edu

Ayoub Daliri is now at the College of Health Solutions, Arizona State University, Tempe.

Nicole Enos is now at Epic Systems Corporation, Madison, WI.

Ashling A. Lupiani is now at the Joint Department of Biomedical Engineering, The University of North Carolina at Chapel Hill and North Carolina State University, Raleigh-Durham.

Editor-in-Chief: Bharath Chandrasekaran

Editor: Kate Bunton

Received April 21, 2020

Revision received July 16, 2020

Accepted August 12, 2020

https://doi.org/10.1044/2020_JSLHR-20-00192

According to current computational models of speech production, sensorimotor control of articulation and voice is governed by feedback and feedforward control systems (Guenther, 1994; Guenther et al., 2006; Houde & Nagarajan, 2011; Parrell et al., 2019). The *feedback controller* uses sensory information (i.e., auditory and somatosensory feedback) to immediately adjust motor commands that are sent to the articulators (i.e., vocal tract) and the larynx. Early in speech development, sensory feedback informs the *feedforward controller*, which guides production of accurate and efficient motor programs. After the feedforward controller is refined, speakers rely less on feedback for motor control. However, when a mismatch between the intended speech or voice output and the actual output is detected, the feedback controller generates an error signal. The detection of this mismatch and potentially the scaling of the error signal depend on the system's *sensory acuity*. Motor commands are quickly adjusted based on the degree of mismatch to correct the output. These responses, which are mediated by the feedback controller and are often referred to as *reflexive responses* (Burnett

Disclosure: The authors have declared that no competing interests existed at the time of publication.

et al., 1997; Hain et al., 2000; Larson & Robin, 2016), occur with sudden artificial perturbations of feedback. When error signals are generated consistently over time, the feedforward controller is also gradually updated to revise subsequent motor programs. Thus, *adaptive responses* are mediated by interactions between the feedforward and feedback control systems and are seen when feedback is consistently perturbed. Previous studies of individuals with typical speech (Burnett et al., 1998, 1997; S. H. Chen et al., 2007; Hain et al., 2000; Houde & Jordan, 1998, 2002; Jones & Munhall, 2000; Liu & Larson, 2007; Purcell & Munhall, 2006; Tourville et al., 2008; Villacorta et al., 2007), individuals with speech disorders (X. Chen et al., 2013; Liu et al., 2012; Mollaei et al., 2016, 2019, 2013), and computational models (Callan et al., 2000; Guenther et al., 2006; Villacorta et al., 2007) have clarified the mechanisms of auditory–motor control of speech. However, while individuals with typical speech generally compensate for perturbations in auditory feedback by opposing the perturbation, many studies have shown variability in response magnitude and direction across individuals (Burnett et al., 1998, 1997; S. H. Chen et al., 2007; Hain et al., 2000; Houde & Jordan, 1998, 2002; Liu & Larson, 2007; Purcell & Munhall, 2006; Tourville et al., 2008; Villacorta et al., 2007). The sources of this variability are currently unknown. Thus, the aim of this study is to assess both reflexive and adaptive auditory–motor control of articulation using first formant (F_1) perturbation and of voice using fundamental frequency (f_o) perturbation within the same sample of individuals with typical speech to further investigate the sources of response variability.

The effects of auditory feedback on speech and voice output have been studied previously using sudden perturbations of F_1 to investigate reflexive articulatory motor control and sudden perturbation of the source spectrum (i.e., f_o and its harmonics) to investigate reflexive vocal motor control. For instance, Tourville et al. (2008) recorded healthy speakers producing monosyllabic words while the F_1 of their auditory feedback was unexpectedly perturbed upward or downward by 30%. They found significant compensatory changes in F_1 in response to the auditory perturbations (i.e., as the F_1 of the auditory feedback increased, the F_1 of the speech output decreased and vice versa), with an average magnitude of about 13% of the applied F_1 perturbation and a range of 4%–26%. The average response latencies were 108–165 ms. This paradigm was similar to those used in prior studies, which applied unexpected upward and/or downward perturbation of the source spectrum during sustained vowel production, although the duration and timing of the perturbations differed (i.e., source spectrum perturbations were generally applied for shorter durations and for multiple instances per trial; see Larson & Robin, 2016, for a review). With 25–30 cents¹ perturbation of the source spectrum, average response magnitudes were 22–30 cents and ranged from

about 3 to 100 cents (Burnett et al., 1998; S. H. Chen et al., 2007). Differences in the magnitude of the responses across perturbation magnitudes within this range were not significant (Burnett et al., 1998; S. H. Chen et al., 2007). The direction of these f_o responses most often opposed the perturbation, although there were also “following” responses, wherein, for example, the f_o of the voice output increased as the f_o of the auditory feedback was increased (Burnett et al., 1998, 1997; S. H. Chen et al., 2007; Liu & Larson, 2007). Average initial response latencies were about 154–228 ms (Burnett et al., 1998, 1997; S. H. Chen et al., 2007). These studies demonstrated that speakers use auditory feedback to quickly correct errors in both articulation and voice. However, there was a high degree of variability in speakers’ response magnitudes and response directions for which the studies were not able to account.

The effects of auditory feedback on adaptive articulatory and vocal motor control have also been studied, using sustained perturbations of F_1 and the source spectrum. Several research groups have investigated the effects of consistent auditory perturbation of vowel formants using different methods (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006; Villacorta et al., 2007). With F_1 perturbation methods similar to Tourville et al. (2008) described above, Villacorta et al. (2007) recorded healthy speakers producing single words in four ordered phases: (a) with the F_1 of their auditory feedback unperturbed, (b) with the F_1 of their auditory feedback gradually perturbed upward or downward across trials, (c) with the F_1 of their auditory feedback maintained at the largest perturbation of $\pm 30\%$, and (d) with the F_1 of the auditory feedback again unperturbed. On average, participants responded across the experiment by progressively opposing the F_1 perturbation. These compensatory F_1 responses were initially seen during the second phase of the experiment when feedback was perturbed. However, participants continued to produce opposing responses into the last phase of the experiment when auditory feedback was returned to normal, demonstrating adaptation to the perturbation. Although the average response was opposing, there was considerable variability, with some participants compensating for the perturbations and others following the perturbation. Using a similar adaptation paradigm, Jones and Munhall (2000) altered the source spectrum of the auditory feedback up to 100 cents, while participants produced sustained vowels across four ordered phases. The results of the study were similar to the F_1 adaptation study findings—as the frequency of the auditory feedback gradually increased or decreased, participants progressively opposed the frequency perturbation. After the perturbation was removed, speakers’ f_o gradually returned toward baseline. The magnitude of these responses was about ± 20 cents relative to the control condition. These findings demonstrated that speakers use auditory feedback to gradually update their feedforward control systems. It is unknown whether or not there was also variability in speakers’ response magnitudes or directions because the study reported only averages of all speakers’ responses regardless of the direction.

¹The cents scale is a logarithmic scale of frequency. Cents can be calculated using the formula $1,200 \times \log_2(f_2/f_1)$, in which f_2 is the perturbed f_o and f_1 is the unperturbed f_o .

Two studies investigated the relation of sudden perturbation of F_1 and sustained perturbation of F_1 in the same sample speakers. Franken et al. (2019) asked native speakers of Dutch to produce the sustained vowel /e/ while the F_1 of their auditory feedback was perturbed upward or downward by 6.7%. These authors found that participants produced smaller responses to sudden perturbations than sustained perturbations, although participants' response magnitudes in the two experiments were not significantly related. Parrell et al. (2017) asked speakers with cerebellar degeneration and healthy older speakers to produce single words while the F_1 of their auditory feedback was perturbed upward or downward by 150 Hz. They found that speakers with cerebellar degeneration produced smaller responses to sustained perturbations in the late phase of exposure and after the perturbation was removed but produced larger responses to sudden perturbations compared to controls. There was no significant relationship between responses to sustained perturbations in the late phase of exposure and responses to sudden perturbations in the last 100 ms of the response for speakers with cerebellar degeneration or controls. These findings suggest that reflexive and adaptive control of F_1 may be driven by distinct mechanisms.

Two studies have investigated both articulatory and vocal auditory–motor control in the same sample of speakers. Max et al. (2003) applied unexpected upward and downward perturbations of the first two formants and the f_0 of the auditory feedback during sustained vowel production. They found opposing responses to formant perturbation and only upward responses to f_0 perturbation, regardless of the direction of the perturbation. Perturbation of the first two formants in this study would have approximated a change in the overall length of the vocal tract rather than an articulatory error (Villacorta et al., 2007). Therefore, the results of this study are expected to differ from those applying only F_1 perturbation. Unexpectedly, the f_0 responses in this study were not consistent with prior studies showing predominant compensatory f_0 responses despite the direction of the perturbation discussed above. It is possible that this is due to f_0 being perturbed while the associated harmonics remained stable, as there was no mention of a frequency perturbation across the full source spectrum. Other methodological differences that could have contributed to this unique finding include different vowel targets (i.e., /E, A, ə/ vs. /a/ in other studies) or different levels of auditory feedback amplification (i.e., +2 dB SPL vs. +10 dB SPL in other studies). Mollaei et al. (2016) also investigated responses to unexpected F_1 and f_0 perturbation in speakers with Parkinson's disease and healthy older controls. They found that individuals with Parkinson's disease had smaller compensatory responses to F_1 perturbation and larger compensatory responses to f_0 perturbation relative to controls, who also exhibited compensatory responses on average. However, the relationship between F_1 and f_0 responses within speakers was not investigated, nor was the relation of the responses to sudden and predictable perturbation of the F_1 and f_0 of the auditory feedback. This

information is essential to characterizing function of both the feedforward and feedback control systems.

In summary, although previous studies of reflexive and adaptive articulatory and vocal control have demonstrated that speakers generally compensate for perturbations of their auditory feedback, these studies also revealed substantial interspeaker variability in both the magnitude and direction of speakers' responses. One factor that might account for the variability in speakers' responses to auditory perturbation is sensory acuity, as this is thought to be important for sensorimotor error detection and correction. However, to our knowledge, only one study has investigated the relation of auditory acuity to F_1 and speech motor control in typical speakers. Villacorta et al. (2007) used an adaptive just noticeable difference (JND) procedure to study auditory acuity to F_1 and adaptive responses to F_1 perturbation in healthy speakers. The study revealed that participants with more sensitive auditory acuity to differences in the F_1 of their own recorded voice signals had larger adaptive responses to predictable perturbation of the F_1 of their auditory feedback. Similarly, Martin et al. (2018) found that participants with more sensitive auditory acuity to pure-tone frequencies and melodies also had larger adaptive responses to predictable perturbation of the F_1 of their auditory feedback, although auditory acuity to F_1 was not assessed in the study. Auditory acuity to pure-tone frequencies and responses to predictable f_0 perturbations have been studied in speakers with Parkinson's disease and healthy age-matched controls (Abur et al., 2018). These measures were not found to be related in either group; however, the control participants had an average age of 64 years (range: 50–77 years) and elevated, although age-appropriate, hearing thresholds. Mollaei et al. (2019) found that individuals with Parkinson's disease had reduced auditory discrimination of F_1 perturbation during a passive listening task and increased auditory discrimination of f_0 perturbation during an active production task relative to control participants, but the association between auditory discrimination thresholds and adaptive and reflexive response magnitudes was not assessed. As such, the relation of auditory acuity to f_0 and adaptive responses to f_0 perturbation have not yet been studied in typical young adult speakers, although auditory acuity to pitch has been studied extensively using pure and complex tones (see Plack & Oxenham, 2005, for a review) and is known to be affected by factors that also influence speech motor control, like musical training (Tervaniemi et al., 2005; Zarate & Zatorre, 2005).

Therefore, the purpose of this study is to explore the relationship between reflexive and adaptive F_1 and f_0 responses and auditory acuity to differences in F_1 and f_0 within the same speakers. Specifically, the aims of this study are to determine (a) if the magnitude of speakers' reflexive and adaptive responses to F_1 and f_0 perturbation are related and (b) if speakers' auditory acuity to F_1 and f_0 predict the magnitude of speakers' responses. A study of this nature could help clarify whether the variability in speakers' responses to F_1 and f_0 perturbation are due to differences in their ability to detect errors in their auditory

feedback and subsequently revise motor commands within both the feedback and feedforward controllers. The results of this exploratory study may also clarify the relation of feedback and feedforward control of articulation and voice in typical speakers and inform future studies of feedback and feedforward control deficits in individuals with motor speech disorders, including Parkinson's disease. Specifically, while previous studies have investigated the differences between reflexive F_1 and f_0 responses between speakers with Parkinson's disease and healthy controls (Mollaei et al., 2016), the relation of auditory acuity to pure-tone frequencies and adaptive f_0 responses in speakers with Parkinson's disease (Abur et al., 2018), and the relation of passive and active F_1 and f_0 auditory acuity in speakers with Parkinson's disease (Mollaei et al., 2019), the relation of reflexive and adaptive F_1 and f_0 responses and passive and active auditory acuity to F_1 and f_0 has not been investigated in speakers with Parkinson's disease. A study of this nature could clarify the underlying deficits in speech and voice motor control in speakers with differing presentations of Parkinson's disease and guide the development of targeted treatment.

Method

Participants

Twenty healthy women between the ages of 18 and 26 years (mean of 20 years) were included in this study.² All participants reported that they were native English speakers; denied speaking a tonal or nasal language; and denied a history of voice, speech, language, cognitive, hearing, and neurological disorders. Volunteers with non-American English accents, persistent glottal fry, or an abnormally low pitch based on a phone screening were excluded from this study. The average Consensus Auditory–Perceptual Evaluation of Voice (Kempster et al., 2009) overall severity rating for the 20 participants was 8/100 (range: 0–30/100) based on the procedures described below. Professional musicians and musicians with graduate-level training were also excluded from this study. Detailed information about the nature and duration of musical experience was obtained for each participant due to the effects of musical experience on auditory attention (e.g., Strait et al., 2010), speech discrimination (e.g., Parbery-Clark et al., 2012), pitch acuity (e.g., Tervaniemi et al., 2005), and auditory–motor control of pitch (Zarate & Zatorre, 2005). Each participant provided written consent prior to data collection in compliance with the Boston University Institutional Review Board.³

²Pilot testing revealed that frequency perturbation induced artifactual amplitude modulation for voices with an f_0 of 130 Hz or below. Thus, males were excluded from this study to maximize the naturalness of the perturbed auditory feedback.

³Four additional participants met the inclusion criteria and were enrolled in this study but were excluded from the final data set due to inability to return for the second day of testing, inability to remain awake throughout the testing session, inability to follow instructions to maintain a steady pitch during sustained vowel productions, or incorrect amplification of auditory feedback during testing.

Procedure

For all procedures, participants were seated in a sound-attenuating booth. Study volunteers underwent a standard hearing screening prior to data collection as described below. A voice screening was also performed to verify that the volunteers had an f_0 higher than 130 Hz. Individuals who passed the hearing screening and the voice screening were invited to participate in the study involving sudden and predictable perturbation of F_1 and f_0 and auditory acuity to F_1 and f_0 experiments. Tasks were completed across two study sessions, each lasting 1–2 hr, to minimize participant fatigue. The order of experiments was designed to minimize the saliency of auditory perturbations early in the experiment. As such, participants completed the following tasks during Session 1: predictable F_1 and f_0 perturbation and “passive” F_1 and f_0 acuity (described below). Audio recordings of the speech protocol were collected either at the end of the first session or at the beginning of the second session, depending on time constraints. The following tasks were completed during Session 2: sudden F_1 and f_0 perturbation and “active” F_1 and f_0 acuity (also described below). All F_1 and f_0 perturbations in this study were in the upward direction to reduce study time and prevent vocal fatigue. The procedures are outlined in Figure 1.

Hearing Screening

A hearing screening was performed with each study volunteer using a GSI 18 audiometer and 3M E-A-RTONE Gold 3A insert earphones. All participants responded to pulsed pure tones at octave intervals between 1000 and 4000 Hz presented at 25 dB HL in both ears based on hearing impairment screening guidelines for adults (American Speech-Language-Hearing Association, 2020).

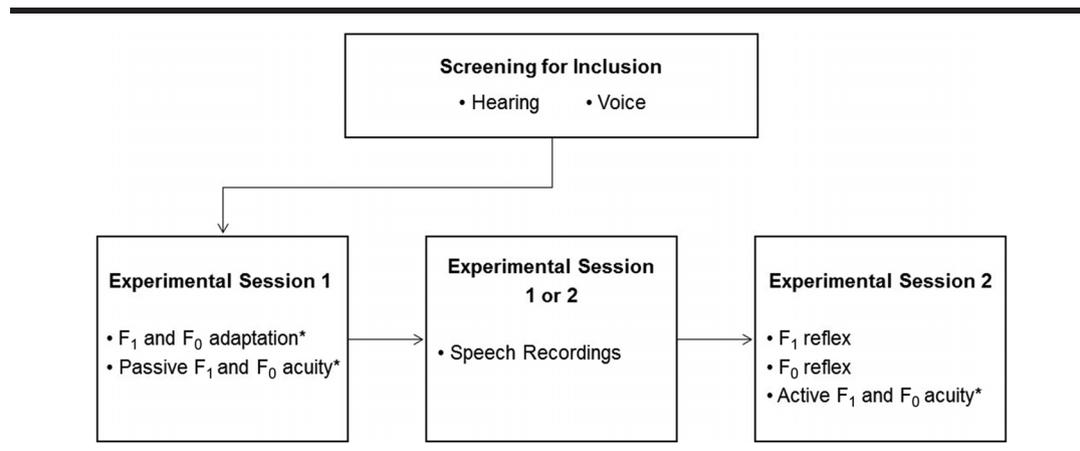
Voice Screening

After the hearing screening, a voice screening was performed to verify that volunteers' f_0 was higher than 130 Hz. An omnidirectional earset microphone (Shure Earset MX153) was positioned 7 cm away and approximately 45° off-axis from the corner of the mouth and was secured to the side of the face using medical tape. Individuals were instructed to produce a sustained /a/ for 2–3 s. The microphone signal was amplified (Behringer XENYX 802 mixer), digitized with a 48000-Hz sampling rate (MOTU MicroBook IIc), routed to a desktop computer (Lenovo ThinkCentre M83), and recorded with a 41000-Hz sampling rate using Praat software (Version 5.4.19; Boersma & Weenink, 2016). The mean f_0 for each vowel production was estimated in Praat.

Data Collection

Equipment. Speech recordings were collected using the same equipment and configuration described above for the voice screenings. In addition, an ASIO4ALL Version 2 sound card driver interfaced with MATLAB (Version 8.1;

Figure 1. Order of screening procedures and experimental tasks. * indicates task order was pseudorandomized and counterbalanced across participants. f_o = fundamental frequency; F_1 = first formant.



MathWorks, 2013) and Audapter 2.1⁴ (Cai et al., 2008; Tourville et al., 2013) software. Speech signals were perturbed in Audapter based on the F_1 and f_o parameters described below for each experiment.⁵ The perturbed and unperturbed output signals for all F_1 and f_o experiments were amplified about 5 dB above the microphone signal by adjusting the output gain in the sound card software (CueMix FX, MOTU MicroBook IIc) and were presented to the participant via insert earphones (Etymotic Research ER-2). These earphones were selected because of their relatively flat frequency response and 90+ dB attenuation of air-conducted noise with deep insertion of foam tips (Etymotic Research 3A/5A) per manufacturer guidelines. The instrumentation was calibrated for each experimental script using a Brüel & Kjær 2 cc Coupler Type 4946, a Brüel & Kjær Type 2250 SPL meter, and a 1000-Hz pure tone played with an Olympus Linear PCM Recorder LS-10. The processing delay from the time of voice onset in the recorded microphone signal to the time of voice onset in the perturbed and unperturbed output signals from the XENYX mixer was 20 ms or less for F_1 perturbation experiments and 45 ms or less for f_o perturbation experiments. Recordings of the microphone input and the Audapter output signals were saved with a sampling frequency of 16000 Hz.

Adaptive responses. Participants' responses to sustained perturbation of their auditory feedback were assessed in two experiments: (a) F_1 adaptation involving a gradual increase of F_1 during production of single words and (b) f_o

adaptation involving a gradual increase of the source spectrum during production of a sustained vowel. The order of these experiments was pseudorandomized and counterbalanced across participants. Prior to these experiments, participants were informed that their speech would be recorded and that they would hear it in the earphones as they spoke. They were not provided with information about the characteristics of the auditory feedback, nor were they provided with instructions for responding to perturbation of auditory feedback.

F_1 adaptation. Participants were instructed to produce a prolonged, steady word for about 1 s when a word appeared on the computer screen. The words “bid,” “tid,” and “hid” were selected as target words for this study because they continued to be real words (i.e., “bed,” “ted,” “head”) when upward perturbation was applied to F_1 . The exact duration of the word was not controlled because vowel duration was not found to affect formant adaptation responses in a prior study (Tourville et al., 2008). Participants were provided with additional verbal and visual cues as needed to produce a comfortable pitch, loudness, and rate during three practice blocks, each consisting of the three target words presented in random order. The auditory feedback for all nine practice trials was unperturbed. The recording time for each trial was 3 s, which began when the target word was presented on the screen and continued for 1 s after the target word was removed from the screen to ensure that auditory feedback was received for the full word production. There was a randomly determined intertrial interval of 1, 1.5, or 2 s to prevent participants from developing a constant rhythm and automatic nature to their productions.

Following the practice trials, participants performed the F_1 adaptation experiment, which consisted of 108 total trials divided into 36 blocks (i.e., the three target words in random order in each block). The blocks spanned four ordered phases: (a) *baseline*—eight blocks with unperturbed auditory feedback, (b) *ramp*—10 blocks with the F_1 of the

⁴Audapter is software for real-time acoustical manipulation of speech. To reduce the computational load for real-time processing in Audapter, speech signals were down-sampled by a factor of 3, resulting in a sampling rate of 16000 Hz and buffer length of 32 samples. Down-sampled signals were then perturbed in Audapter as specified below for each experiment.

⁵For the f_o adaptation experiment, signal amplification was adjusted in MATLAB using an amplitude correction factor of 1.2 for perturbed trials and 0.8 for unperturbed trials to reduce the changes in signal amplitude related to manipulation in Audapter.

auditory feedback gradually increased by 1.03% per trial relative to the produced speech signal (Ramp Trial 1 = 0% perturbation, Ramp Trial 30 = 30% perturbation), (c) *hold*—10 blocks with the F_1 of the auditory feedback maintained at the maximum perturbation of 30% relative to the produced speech signal, and (d) *after effect*—eight blocks with unperturbed auditory feedback. The total number of trials and the number of trials in each phase were sufficient for generating an adaptive response while reducing fatigue based on pilot testing. The total number of trials was higher than in recent F_1 experiments (Franken et al., 2019; Parrell et al., 2017), but lower than an earlier study (Villacorta et al., 2007). The duration of the recordings and the inter-trial intervals for the 108 experimental trials were identical to those for the nine practice trials described above.

f_0 adaptation. Participants were instructed to produce a steady, sustained /a/ when the visual cue “aaa” appeared on the computer screen for 2 s. They were given additional verbal and visual cues as needed to produce a comfortable, steady pitch and loudness and to achieve the 2-s target duration during nine practice trials with unperturbed auditory feedback. The recording time for each trial was 3 s, which began when the visual cue appeared on the screen and continued for 1 s after the cue was removed from the screen. The intertrial interval was 1, 2, or 3 s, with the duration randomly determined. These intervals allowed for slightly longer rest breaks following the 2-s vowel production time compared with the 1-s word production time in the F_1 adaptation.

After completion of the practice trials, participants performed the f_0 adaptation experiment in two conditions (i.e., control and shift-up) with the order pseudo-randomized and counterbalanced across participants. The control condition included 108 trials of sustained /a/ production with unperturbed auditory feedback. The shift-up condition included 108 trials divided into four ordered phases: (a) *baseline*—24 trials with unperturbed auditory feedback, (b) *ramp*—30 trials with the source spectrum of the auditory feedback gradually increased by 3.4 cents relative to the produced voice signal (Ramp Trial 1 = 0 cents perturbation, Ramp Trial 30 = 100 cents perturbation), (c) *hold*—30 trials with the auditory feedback maintained at the maximum perturbation of 100 cents relative to the produced voice signal, and (d) *after effect*—24 trials with unperturbed auditory feedback. The duration of the recordings and the intertrial intervals for the 108 experimental trials were identical to those for the nine practice trials described above.

Reflexive responses. Participants’ responses to sudden perturbation of their auditory feedback were assessed in two experiments: (a) F_1 reflex experiment involving unpredictable upward perturbation of F_1 during production of single words and (b) f_0 reflex experiment involving unpredictable upward perturbation of the voice spectrum (i.e., f_0 and harmonics) during production of a sustained vowel. All participants performed the F_1 reflex before the f_0 reflex due to the more salient perturbation of auditory feedback in the f_0 reflex experiment.

F_1 reflex. Participants were instructed to produce a prolonged, steady word that lasted about 1 s whenever a word appeared on the computer screen. The words “bid,” “tid,” and “hid” were selected as target words to align with the F_1 adaptation experiment. Participants were not informed that they would hear a change in their auditory feedback. Participants completed one practice block, which consisted of the three target words produced 3 times in random order. The F_1 of the auditory feedback was increased by 30% (Tourville et al., 2008) relative to the produced speech signal for two of the nine trials within the block. The F_1 perturbation began at voicing onset and persisted until the end of the trial (Houde & Jordan, 1998). Perturbation was not applied to the first or last trial in a block, and there were at least three unperturbed trials between each perturbed trial to limit adaptation. Otherwise, perturbation was randomly applied to two trials within the block.

Following the practice trials, participants performed the F_1 reflex experiment, which consisted of 12 blocks of the three target words produced 3 times each in a random order (108 total trials). The frequency and magnitude of formant perturbation were identical to those described in the practice block. In total, the auditory feedback was perturbed in 24 trials and unperturbed in 84 trials, which was below the 2:6 perturbed-to-unperturbed ratio in Tourville et al. (2008).

f_0 reflex. Participants were instructed to sustain the vowel /a/ at a comfortable pitch and loudness when the visual cue “aaa” appeared on the computer screen for 2 s. They were informed that they would hear a change in their pitch during some trials to avoid startling participants with the obvious, abrupt changes in their auditory feedback. However, they were not given instructions for responding to the perturbation. Participants completed one practice block consisting of nine trials. The spectrum of the auditory feedback was increased by 100 cents relative to the produced voice signal (Burnett et al., 1998) for two of the nine trials within the block. The onset of the perturbation in this experiment occurred randomly between 0.05 and 1 s following voicing onset to allow the voice to stabilize before the perturbation was applied. The perturbation persisted until the end of phonation (Burnett et al., 1997). Like the F_1 reflex experiment, perturbation was not applied to the first or last trial in a block, and there were at least three unperturbed trials between each perturbed trial to limit potential adaptation.

After the practice block, participants performed the f_0 reflex experiment, which included 12 blocks of sustained /a/ produced for nine trials (108 total trials). The frequency, magnitude, and onset of the perturbation were identical to those described in the practice block. In total, the auditory feedback was perturbed in 24 trials and unperturbed in 84 trials.

Auditory acuity. Participants’ acuity to F_1 and f_0 was assessed using four JND experiments described below. All JND experiments implemented an adaptive one-up, two-down staircase procedure (Levitt, 1971), with a 1:1 up/down

ratio to obtain a target threshold of 70.71% (García-Pérez, 1998). The step sizes, minimum and maximum F_1 or f_0 differences, and number of reversals were optimized for efficiency and consistency of JND estimation during pilot testing for each experiment.

Passive. Two “passive” JND experiments were conducted, wherein word or vowel productions were recorded prior to a listening experiment. These samples were subsequently changed during listening experiments involving a pair comparison task. This task was included to assess participants’ auditory acuity to perturbation of their own productions while the speech motor control system was not engaged in generating speech, similar to previous studies (Mollaei et al., 2019; Villacorta et al., 2007). For the two perceptual experiments, the intensity of the auditory stimuli was normalized and presented at 75 dB SPL to all listeners. The passive experiments were always performed after the adaptation experiments and before the F_1 reflex experiment. The order of passive F_1 and f_0 experiments was pseudorandomized and counterbalanced across participants.

F_1 : Participants were instructed to produce a prolonged, steady word for about 1 s whenever the word “bid,” “bed,” or “bad” appeared on the screen. The words “bed” and “bad” were selected to prevent the participant from knowing that “bid” would be the target word used in the listening experiment. Their productions were recorded in Audapter, and the production with the median F_1 was selected for the subsequent listening study. Participants were presented with a pair of the selected “bid” recordings with a 500-ms interstimulus interval. One stimulus in the pair served as the reference stimulus and had an upward F_1 perturbation of 1%. The other stimulus either had an upward change of 1% for catch trials or had an upward change with the magnitude based on the adaptive procedure. The 1% change was applied to the reference stimuli in an attempt to process all signals in a similar way in Audapter, thus minimizing any acoustical differences aside from the magnitude of F_1 perturbation. The change in F_1 was applied for the duration of the word. Twenty percent of the trials were catch trials. The participants’ responses for these trials were not used in the logic of the adaptive procedure but were used to assess participants’ attention to the task. The initial F_1 change applied in the adaptive procedure was a 40% increase. For each trial, participants were asked if the vowel sounds in the two words were the same or different. The trial advanced as soon as the participant’s response was entered by the investigator. When the participant responded that the vowel sounds were different during two consecutive trials with F_1 change based on the adaptive procedure, the F_1 change decreased by a fixed step size of 3% of the baseline in the subsequent trial. When the participant responded that the vowel sounds were the same during a trial with F_1 change in the adaptive procedure, the F_1 increased by a step size of 3% of the baseline in the subsequent trial. The maximum F_1 change was 60%, and the minimum F_1 change was 1%. Prior to the full experiment, participants produced the target words and performed

nine practice listening trials to familiarize them with the speech production and listening tasks. The full experiment included either 10 reversals (i.e., changes in the direction of the staircase) or 60 total trials, whichever was shorter.

f_0 : Participants produced the vowel /a/ for 3 s when a visual prompt appeared on the computer screen. Their production was recorded in Audapter, and a waveform of the recording was inspected for signal quality. The signal was rerecorded if necessary. A 1-s segment of the recording was manually selected from the midportion of the production with a relatively steady amplitude. From this signal, a 500-ms segment was automatically selected and a 50-ms cosine squared ramp was applied to attenuate abrupt signal onset and offset. This signal was then used in a pair comparison practice set and full listening experiment similar to the formant procedure described above. The only differences were that the reference stimulus had an upward change in the source spectrum by 0.01 cents, the initial upward change in the adaptive procedure was 50 cents, the step size was 4 cents, the maximum change was 200 cents, and the minimum change was 0.01 cents.

Active. Two “active” JND experiments were conducted, wherein speech or voice samples were recorded and perturbed online during a combined production and perception experiment. This task was included to assess participants’ auditory acuity to perturbation of their own productions while the system was engaged in generating speech output, similar to Mollaei et al. (2019). The order of F_1 and f_0 experiments was pseudorandomized and counterbalanced across participants.

F_1 : Participants were instructed to produce a prolonged “bid” when the word appeared on the screen. They were informed that their speech would be recorded and that they would hear it in their earphones as they spoke. They were also informed that the “bid” that they would hear would sometimes have a different vowel sound than the “bid” that they produced. They were instructed to judge whether the vowel that they heard sounded the *same* as the vowel that they produced or *different* from it. The listening experiment was performed in a similar manner to the passive formant JND, except that the stimuli were not presented in pairs. In addition, the initial perturbation was a 65% upward perturbation of F_1 , the step size was 4%, the maximum perturbation was 100%, and the minimum perturbation was 0%. The catch trial frequency was increased to 50% in this experiment in an effort to prevent adaptation to the perturbation; these trials again were not used in the logic of the adaptive procedure. Participants performed nine practice trials prior to the full experiment, which included either 15 reversals or 75 trials, whichever was shorter.

f_0 : Participants were asked to produce the vowel /a/ for 2 s when a prompt appeared on the computer screen. They were told that their speech would be recorded and that they would hear it in their earphones as they spoke. They were also told that the /a/ that they would hear would sometimes have a different pitch than the /a/ that they produced. They were instructed to decide if the pitch that they heard sounded the *same* as the pitch that they produced

or *different* from it. The practice set and full listening experiment procedures were consistent with the active formant JND. The initial perturbation was a 75-cent upward perturbation to the source spectrum, the step size was 6 cents, the maximum perturbation was 400 cents, and the minimum perturbation was 0 cents.

Audio recordings. The insert earphones were removed. Participants were instructed to read *The Rainbow Passage* (Fairbanks, 1960) using their normal speaking voice. Audio recordings were collected using the same earset microphone as the perturbation experiments (Shure Earset MX153) in SONAR Artist (Cakewalk, Inc.), with a sampling rate of 44100 Hz.

Data Analysis

Adaptive responses

F_1 adaptation. F_1 estimation was performed online in Audapter using linear predictive coding (LPC), with an LPC order of 16 coefficients. Custom-written MATLAB (Version 8.2; MathWorks, 2013) scripts were used off-line to analyze the mean F_1 within a 50- to 120-ms window of the identified word segment for each of the single words produced during the experiment.⁶ The mean F_1 was then calculated for the words produced during the 24 baseline trials, and the percentage of deviation of the mean F_1 from the baseline mean was determined for all trials. The normalized mean F_1 in three-trial blocks for the hold phase was calculated and used in the statistical analyses described below.

f_0 adaptation. Custom-written MATLAB and Praat (Versions 5.3–6.0; Boersma & Weenink, 2016) scripts were used to analyze the mean f_0 within a window of 50–120 ms after vowel onset for each vowel produced in the experimental trials. Estimates of the f_0 were obtained via an autocorrelation method in Praat.⁷ The mean f_0 was then

⁶This analysis window was chosen in order to include the feedforward plan prior to generating a reflexive response (< 120 ms) while excluding initial fluctuations due to biomechanical variability of the vocal folds and formant transitions (< 50 ms). The longest duration of the F_1 tracked for each signal was selected as the speech segment for analysis. The selection was plotted with the F_1 values across time, and a wideband spectrogram for visual verification that the automated selection was appropriate. When it was not, a segment of the word that started at onset of the F_1 trace was manually selected. The mean F_1 for the analysis window (50–120 ms) was calculated. When a segment that was manually selected contained any values of 0 related to intermittent glottal fry, these values were excluded from the mean.

⁷Default settings within the Praat interface were used, except for the voicing threshold, which was increased to 0.9 to eliminate environmental noise and f_0 instability at phonatory onsets and offsets. Each trial was also visually inspected in Praat using these settings. When f_0 tracking appeared to be inconsistent or inaccurate across trials, the voicing threshold and pitch range settings were adjusted to optimize f_0 estimation, and the automated analysis was performed again using these optimized settings. When f_0 tracking continued to be problematic for individual trials, the voicing threshold and pitch range were optimized for the trial, and the mean f_0 was manually obtained. When the f_0 was automatically tracked during nonspeech noise (e.g., throat clearing) or false starts, the mean f_0 was manually obtained for the portion of the signal including the sustained vowel only.

calculated for the 24 baseline trials, and the mean f_0 for each trial was converted to cents relative to the baseline mean.

The shift-up condition was then normalized to the control condition by subtracting the mean f_0 in cents for each control trial from the associated mean f_0 in cents for each shift-up trial. This normalization was performed to account for any pattern of change in f_0 across the experiment that was not related to the change in auditory feedback. The normalized mean f_0 in three-trial blocks for the hold phase was used in the statistical analyses.

Reflexive responses

F_1 . F_1 estimation was performed online in Audapter using the LPC parameters described above. Custom-written MATLAB scripts were used to obtain the F_1 every 2 ms starting at the onset of the F_1 for each single word and extending 500 ms to capture the F_1 across the initial consonant, formant transition, and vowel.⁸ F_1 traces for unperturbed trials that were not within two trials after a perturbed trial were time-aligned and averaged on a sample-by-sample basis for each word to serve as the baseline. The F_1 trace for each perturbed trial was then normalized to the baseline F_1 trace for the associated word by subtracting the F_1 of the baseline from the F_1 of the perturbed trial on a sample-by-sample basis. The normalized mean F_1 within a post-perturbation window of 120–240 ms was used in the statistical analyses. This window was selected to capture reflexive F_1 responses based on the timing reported by Tourville et al. (2008). The analysis window was selected to align with the f_0 analysis described below and to avoid capturing responses that were unlikely to be generated during typical running speech production.

f_0 . Custom-written MATLAB and Praat scripts were used to analyze the f_0 every 10 ms, beginning 100 ms prior to the perturbation (Behroozmand et al., 2012) and extending 850 ms after the perturbation for each vowel.⁹ Praat settings and f_0 estimation procedures were the same as those described above for the f_0 adaptation experiment, except that when the f_0 could not be obtained for any sample during the analysis period, the trial was excluded. Each f_0 value within the analysis window was converted to cents relative to the mean f_0 of the baseline period. Trials with baseline variability exceeding ± 15 cents were excluded from further analysis (Behroozmand et al., 2012). The normalized mean f_0 within a postperturbation window of 120–240 ms was used in the statistical analyses. This window was selected to be consistent with the analysis window used for the F_1 reflex analyses and to capture primary reflexive f_0 responses while avoiding inclusion of secondary,

⁸The F_1 trace was plotted with a wideband spectrogram for visual verification that the automated identification of the onset of the word was accurate. When the automated selection did not begin at the onset of the F_1 series for the target word, the onset was manually selected.

⁹The perturbation onset time was identified by finding the maximum change in f_0 from one sample to the next in the perturbed signal. The time point of the sample with the higher f_0 was then associated with the time point in the microphone signal.

longer latency responses to sustained f_0 perturbation (Burnett et al., 1998, 1997; Hain et al., 2000).

Auditory acuity. For all JND experiments, the threshold was estimated by calculating the mean F_1 or f_0 difference across the last four reversals in the experiment.

Audio recordings. A certified speech-language pathologist rated the recordings of The Rainbow Passage (Fairbanks, 1960) for each participant using the Consensus Auditory-Perceptual Evaluation of Voice (Kempster et al., 2009). The audio recordings were presented in random order via Sennheiser 280 Pro HD headphones at a comfortable volume, and the speech-language pathologist was blinded to the identity of each participant.

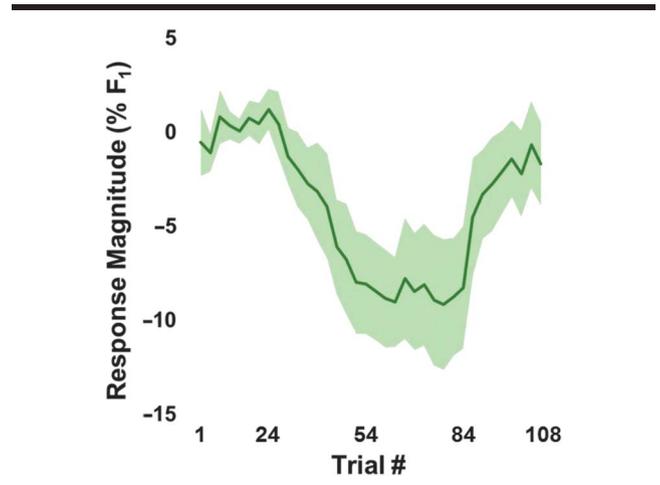
Statistical Analyses

The data were analyzed using two Pearson's correlation analyses to determine if the adaptive F_1 responses were correlated with the reflexive F_1 responses and if the adaptive f_0 responses were correlated with the reflexive f_0 responses. An adjustment for multiple comparisons was not applied due to the exploratory nature of the study. Therefore, a significance level of $p < .05$ was used to interpret the correlation analysis. After determining that the passive and active JND scores were not correlated using two Pearson's correlation analyses, the adaptive and reflexive F_1 responses were modeled using a multivariate general linear model with potential predictor variables representing the passive and active F_1 JND scores. The adaptive and reflexive f_0 responses were also modeled using a multivariate general linear model with potential predictor variables representing the passive and active f_0 JND scores. A significance level of $p < .05$ was used to interpret the models. All statistical analyses were performed using SPSS (Version 25; IBM).

Results

On average, participants exhibited adaptive compensatory responses to the predictable F_1 (see Figure 2) and f_0 (see Figure 3) perturbations. During the hold phase, the average adaptive F_1 response was -8.6% ($SD = 6.2\%$). One participant produced an average following F_1 response of 2.2% , and another participant produced an average F_1 nonresponse of -0.3% . All other participants produced average compensatory F_1 responses between -18% and -2% . The average adaptive f_0 response was -61.2 cents ($SD = 79.0$ cents). Four participants produced average following f_0 responses between 16.1 and 55.2 cents, and one participant produced an average f_0 nonresponse of 0.9 cents. All other participants produced f_0 compensatory responses between -263.2 and -6.0 cents. On average, participants also produced reflexive compensatory responses to the sudden F_1 (see Figure 4) and f_0 (see Figure 5) perturbations. The average reflexive F_1 response was -0.6% ($SD = 1.3\%$). Five participants produced following F_1 responses between 0.4% and 2.4% . All other participants produced compensatory F_1 responses between -0.2% and -3.3% . The average reflexive f_0 response was -7.1 cents ($SD = 6.0$ cents). One participant

Figure 2. Average adaptive response to gradual perturbation of the first formant (F_1), with the shading representing the 95% confidence intervals. The baseline phase included Trials 1–24, ramp included Trials 25–54, hold included Trials 55–84, and after effect included Trials 85–108. Responses were measured in the 50- to 120-ms window for each trial, normalized to the baseline mean, and averaged over three-trial blocks.



produced a following f_0 response of 9.9 cents. All other participants produced compensatory f_0 responses between -0.2 and -17.9 cents. Estimates of auditory acuity in the passive and active JND experiments are presented for F_1 in Figure 6 and f_0 in Figure 7. The average passive F_1 JND score was 15% ($SD = 9\%$), and the average active F_1 JND score was 41% ($SD = 23\%$). The average passive f_0 JND score was 37 cents ($SD = 14$ cents), and the average active f_0 JND score was 39 cents ($SD = 36$ cents).

Figure 3. Average adaptive response to predictable perturbation of the fundamental frequency, with the shading representing the 95% confidence intervals. The baseline phase included Trials 1–24, ramp included Trials 25–54, hold included Trials 55–84, and after effect included Trials 85–108. Responses were measured in the 50- to 120-ms window for each trial, normalized to the control condition, and averaged over three-trial blocks.

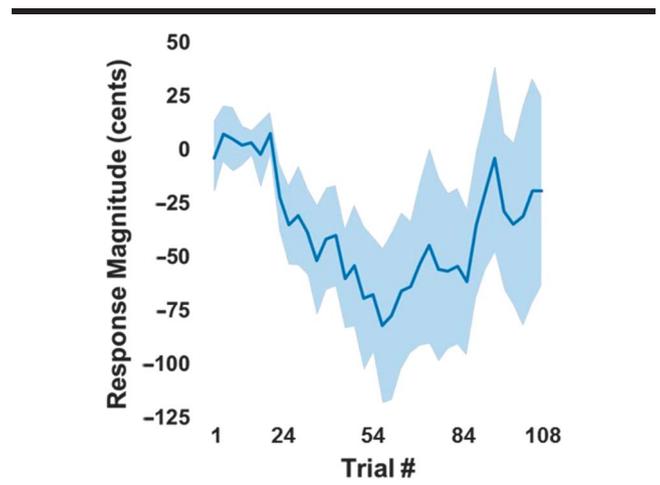
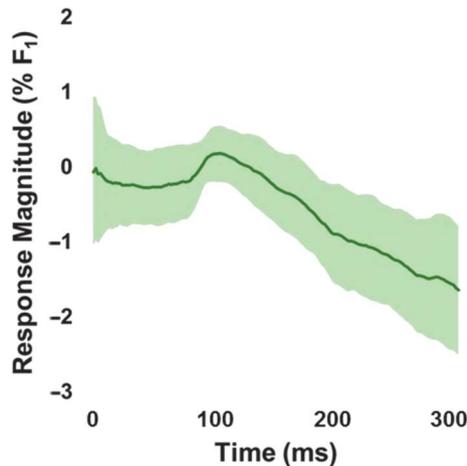


Figure 4. Average reflexive response to sudden perturbation of the first formant (F_1), with the shading representing the 95% confidence intervals. Responses were measured in the 120- to 240-ms window after the perturbation onset at 0 ms for each trial and normalized to the mean of unperturbed trials of the same word.



The average adaptive F_1 responses were correlated with the average reflexive F_1 responses ($r = .468, p = .038$; see Figure 8). The average adaptive f_0 responses were not correlated with the reflexive f_0 responses ($r = .346, p = .135$; see Figure 9). The passive and active F_1 JND scores were not correlated ($r = .281, p = .234$), nor were the passive and active f_0 JND scores ($r = .094, p = .692$). The model of F_1 responses and JND scores was significant for reflexive responses ($R^2 = .437, p = .008$) but not for adaptive responses ($R^2 = .145, p = .263$). Reflexive F_1 responses were significantly

Figure 5. Average reflexive response to sudden perturbation of the fundamental frequency, with the shading representing the 95% confidence intervals. Responses were measured in the 120- to 240-ms analysis window after the perturbation onset at 0 ms for each trial and normalized to the mean of the -100- to 0-ms preperturbation window.

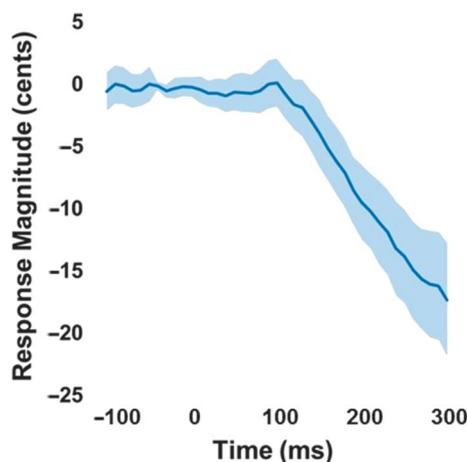
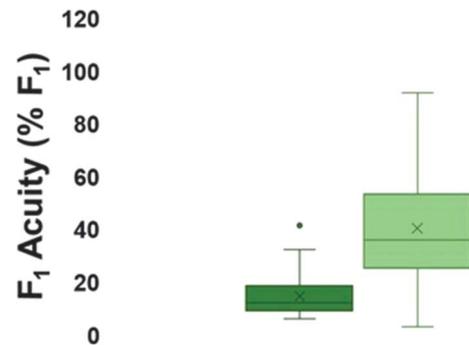


Figure 6. Box-and-whisker plots of first formant (F_1) acuity for the passive (left) and active (right) just noticeable difference tasks, with the minimum, median, and maximum acuity represented by the horizontal lines and the mean represented by \times .



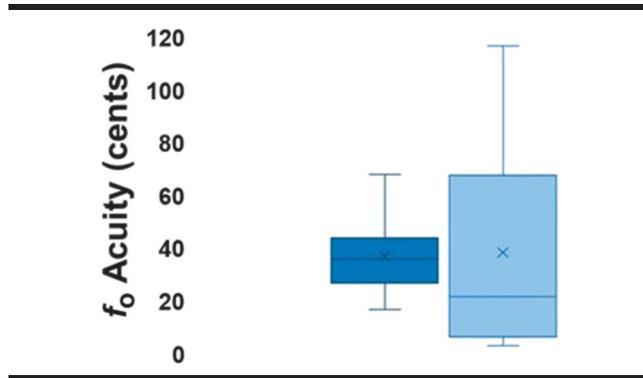
associated with passive F_1 JND, such that larger compensatory responses were associated with lower JND scores ($p = .012$). See Table 1 for further details. The model of f_0 responses and JND scores was not significant for reflexive responses ($R^2 = .008, p = .938$) or adaptive responses ($R^2 = .175, p = .195$; see Table 2 for further details).¹⁰

Discussion

The results of this study revealed that typical speakers' responses to gradual, predictable perturbation of F_1 were related to their responses to sudden, unpredictable perturbation of F_1 , in contrast to previous studies (Franken et al., 2019; Parrell et al., 2017). Furthermore, typical speakers' responses to sudden perturbation of F_1 were associated with their auditory acuity to off-line perturbation of their F_1 . That is, speakers with better auditory acuity to F_1 (lower passive F_1 JND scores) produced larger reflexive responses to sudden perturbations of F_1 . The relationships found between reflexive and adaptive F_1 responses and between reflexive F_1 responses and auditory acuity to F_1 should be interpreted with caution, as correction for multiple comparisons would have negated the significance of the relationships. Unexpectedly, typical speakers' responses to gradual perturbation of F_1 were not associated with their auditory acuity to F_1 , in contrast to the previous findings by Villacorta et al. (2007). In addition, typical speakers' responses to predictable and sudden perturbation of f_0 were not related to each other, nor were their responses related to either measure of auditory acuity to f_0 . These findings may indicate that different gains are applied to corrective changes within the feedback and feedforward control systems for

¹⁰For the f_0 adaptation experiment, one participant's average hold response was based on nine blocks of three trials rather than 10 blocks due to equipment failure. For the f_0 reflex experiment, 17 participants' averages were based on 13–23 trials due to baseline variability exceeding ± 15 cents or due to missing data in the analysis window related to f_0 estimation issues, glottal fry, or short productions.

Figure 7. Box-and-whisker plots of fundamental frequency (f_0) acuity for the passive (left) and active (right) just noticeable difference tasks, with the minimum, median, and maximum acuity represented by the horizontal lines and the mean represented by x .



articulatory and voice error correction based on auditory feedback.

According to contemporary models of auditory–motor control of speech (Guenther, 1994; Guenther et al., 2006; Houde & Nagarajan, 2011; Parrell et al., 2019), a difference between a speaker’s desired auditory signal and actual auditory signal conveyed by their auditory feedback necessitates both feedback (reflexive) and feedforward (adaptive) responses. Thus, questions arise about the nature of shared circuitry in generating these responses. Model formulations that posit more shared circuitry might predict that, given varying gains for a specific speaker, feedback responses would be related to feedforward responses. The finding in this study that participants’ responses to predictable

Figure 8. Relationship between average adaptive responses to gradual perturbation of first formant (F_1) and average reflexive responses to sudden perturbation of F_1 for each participant ($r = .468$, $p = .038$).

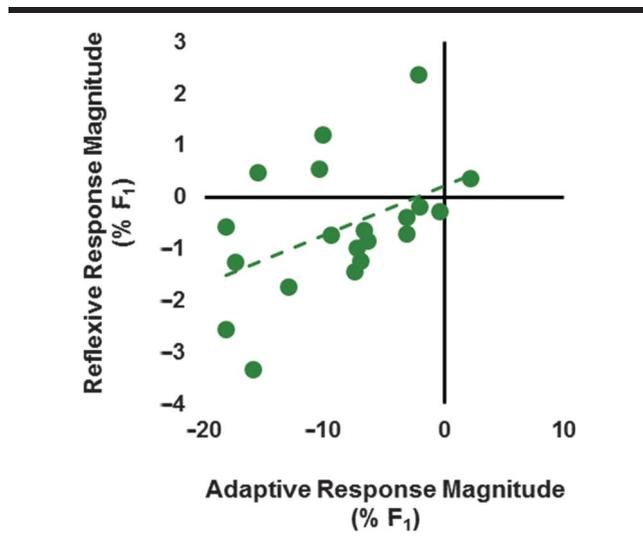
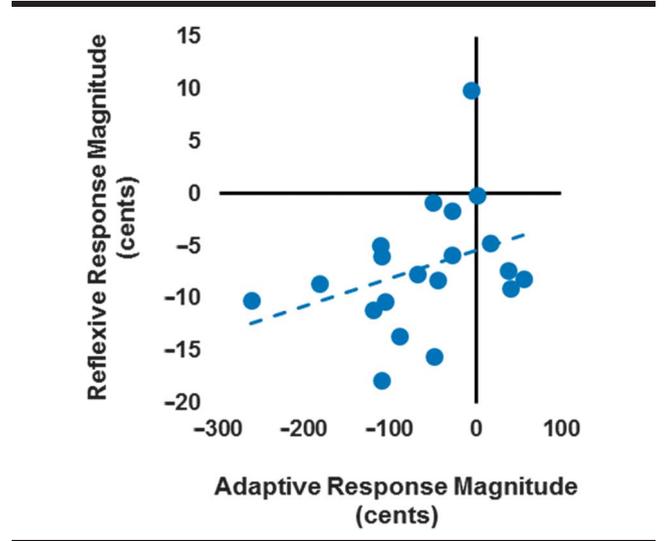


Figure 9. Relationship between average adaptive responses to sudden perturbation of fundamental frequency (f_0) and average reflexive responses to sudden perturbation of f_0 for each participant ($r = .346$, $p = .135$).



and sudden perturbations of F_1 were related is consistent with this idea, whereas the finding that participants’ responses to predictable and sudden perturbations of f_0 were not related is more consistent with dissociation of this control.

Disparate relationships between reflexive and adaptive responses to F_1 (articulatory) and f_0 (voice) parameters may be due to the inherent differences between these two types of speech motor control. Articulatory parameters such as F_1 are used to convey phonemic identity and change rapidly during speech, whereas vocal parameters such as f_0 primarily affect prosody and generally change more slowly. These differences in timing and perhaps control are supported by prior work. Abrupt removal of auditory feedback was examined in individuals with cochlear implants, switching their implant microphones off and on over a single experimental session (Perkell et al., 2007). The authors found that so-called “postural” measures of sound pressure level, duration, and f_0 were heavily influenced by the presence or absence of feedback. In contrast, so-called “contrast measures” of vowel separation and sibilant separation did not change substantially when access to auditory feedback was blocked. This work suggested that vocal parameters are more heavily influenced by auditory feedback control mechanisms than articulatory parameters. This is likely because auditory feedback is too slow for online control of articulatory movements, which must rely more heavily on feedforward control mechanisms (Guenther et al., 2006; Perkell, 2012). Thus, it is perhaps unsurprising that the relationships between articulatory and voice reflexive and adaptive responses differed in this study.

The relationships between reflexive and adaptive control of articulation and voice require further investigation in larger samples of speakers, particularly with regard

Table 1. Results of multivariate general linear model of the first formant (F_1) responses.

Dependent variable	Predictor variable	Coefficient (β)	SE	95% CI	η_p^2	p	Observed power
Reflexive F_1	Intercept	-2.305	0.537	[-3.437, -1.172]	.520	< .001	.981
	Passive F_1 JND	7.512	2.688	[1.840, 13.184]	.315	.012	.750
	Active F_1 JND	1.497	1.034	[-0.685, 3.678]	.110	.166	.277
Adaptive F_1	Intercept	-9.798	3.269	[-16.695, -2.900]	.346	.008	.806
	Passive F_1 JND	26.050	16.374	[-8.495, 60.595]	.130	.130	.324
	Active F_1 JND	-6.432	6.298	[-19.719, 6.856]	.058	.321	.161

Note. SE = standard error; CI = confidence interval; JND = just noticeable difference.

to the relationship between reflexive and adaptive control of F_1 and f_o . Although the current study indicates that there is a relationship between these control mechanisms for F_1 , other studies did not show a significant relationship. Specifically, Franken et al. (2019) found no correlation between reflexive and adaptive responses to F_1 perturbation with 26 speakers of Dutch ($r = .093$, $p = .53$). Similarly, Parrell et al. (2017) found no correlation between the same variables with 14 older controls with age-related hearing loss ($r = .30$, $p = .29$). While previous studies have not investigated the relationship between reflexive and adaptive control of f_o , the current study indicates that the relationship between these control mechanisms is not significant. This finding may have been related to the pattern that participants who closely compensated for the predictable f_o perturbations (i.e., those who produced -100 cent adaptive responses to +100 cent perturbations) had a wide range of reflexive f_o response magnitudes (i.e., -17.9 to -5.0 cents), whereas other participants showed a more direct, positive relationship between adaptive and reflexive f_o response magnitudes. Alternatively, the nonsignificant finding may have been driven by one participant who had the largest reflexive f_o response (i.e., 9.9 cents) and a relatively small adaptive f_o response (i.e., -6.0 cents). Although there was nothing obviously different about this participant's background or performance on other tasks, the participant's data could have masked a significant relationship between reflexive and adaptive control of f_o in other participants, thereby obscuring similarities between articulatory and voice motor control mechanisms.

In addition to needing larger samples of typical speakers for adequate power to detect relationships, future

studies should investigate the analysis windows and blocks that optimize assessment of reflexive and adaptive responses to both F_1 and f_o perturbation. Franken et al. (2019) compared participants' adaptive response magnitudes in a 50- to 150-ms window after speech onset to reflexive response magnitudes in a 1,000- to 1,500-ms window after speech onset, whereas Parrell et al. (2017) used analysis windows of 50–100 ms for adaptive responses in the last 10 trials of the hold phase and 300–400 ms for reflexive responses. Finally, sensory preference should be considered in future studies because this factor may contribute to differences in F_1 and f_o response magnitudes across participants and across studies. Although previous studies have indicated that more speakers rely on auditory feedback than somatosensory feedback for F_1 error correction, while some rely on both (Lametti et al., 2012), it is possible that speakers weigh auditory and somatosensory feedback differently for f_o error correction. Preferential use of somatosensory over auditory information could also contribute to the different patterns seen in articulatory and voice error correction in the current study.

Only the reflexive responses to F_1 were significantly related to participants' corresponding auditory acuity. Participants' responses to predictable and sudden perturbation of f_o were unrelated to their auditory acuity to f_o . In contrast, Villacorta et al. (2007) found a moderate significant relationship between participants' responses to predictable perturbation of F_1 and auditory acuity to F_1 ($R^2 = .312$). Likewise, in children, auditory acuity to f_o has been shown to be related to both reflexive and adaptive responses to f_o : Children with less sensitive auditory acuity to f_o showed significantly larger reflexive and smaller adaptive responses

Table 2. Results of multivariate general linear model of fundamental frequency (f_o) responses.

Dependent variable	Predictor variable	Coefficient (β)	SE	95% CI	η_p^2	p	Observed power
Reflexive f_o	Intercept	-6.208	4.311	[-15.303, 2.888]	.109	.168	.274
	Passive f_o JND	-3.272	10.549	[-25.529, 18.985]	.006	.760	.060
	Active f_o JND	0.855	4.099	[-7.793, 9.504]	.003	.837	.054
Adaptive f_o	Intercept	-134.376	51.475	[-242.979, -25.772]	.286	.018	.692
	Passive f_o JND	231.300	125.967	[-34.467, 497.067]	.166	.084	.410
	Active f_o JND	-32.201	48.948	[-135.474, 71.071]	.025	.519	.095

Note. SE = standard error; CI = confidence interval; JND = just noticeable difference.

(Heller Murray & Stepp, 2020). The lack of significant associations between responses to perturbations and auditory acuity in this study may be due to insufficient power to detect a relationship or may be a function of the narrow range of responses. In Villacorta et al., adaptive responses varied widely, including several participants with “following” responses. Similarly, the responses from children in Heller Murray and Stepp (2020) were heterogeneous, far more so than the adults studied. It may be that relationships exist between auditory acuity and corresponding responses to auditory perturbations, but that the present sample of speakers with typical voice and speech function did not capture sufficient variability to detect it.

Substantial variability was seen in participants’ active JND scores relative to passive JND scores for both F_1 and f_o auditory acuity, but more so for f_o acuity. The pattern of average JND scores, however, was consistent across the passive and active F_1 and f_o acuity experiments; specifically, participants’ auditory acuity was more sensitive (corresponding to a lower JND score) for the passive experiments than the active experiments. These results indicate that auditory acuity to F_1 and f_o may be suppressed during active speech and voice production relative to passive listening. Because passive and active F_1 JNDs provided a better model fit when combined, both tasks may have some predictive value for reflexive F_1 responses and should be investigated further. It should be noted that differences between the passive and active auditory acuity experiments may reflect differences not only in auditory acuity but also in cognitive function. That is, the pair comparison task in the passive listening experiment placed a demand on working memory that the active listening experiment did not. Furthermore, both the passive and active auditory acuity experiments required participants to make a behavioral choice, which may not reflect how the automatic auditory–motor control system detects differences between the intended output and the perceived output. The finding in the current study that passive acuity to F_1 was more sensitive than active acuity to F_1 is consistent with data from healthy control speakers in the aforementioned study of auditory–motor control in Parkinson’s disease (Mollaei et al., 2019). In that study, F_1 discrimination was also more sensitive in passive experiments than active experiments for healthy control speakers. Although, in the same study, f_o discrimination was similar in both passive and active experiments, unlike the current experiment. Methodological differences may have contributed to these inconsistent findings. Mollaei et al. (2019) applied F_1 perturbations of 15% and 30% and f_o perturbations of 25, 50, and 100 cents in their discrimination experiments and reported discrimination results that were collapsed across the two F_1 perturbation magnitudes and the three f_o perturbation magnitudes. As such, passive f_o JND scores of 37 cents and active f_o JND scores of 39 cents seen in the current study may have been obscured if collapsed. Furthermore, participant characteristics may have contributed to differences, including age, which was higher in Mollaei et al. (2019) than in the current study, and compensation to auditory

perturbation, which appeared to be larger for f_o perturbation in Mollaei et al. (2019) than in the current study. As the magnitude of reflexive responses produced during active JND experiments could influence JND scores, future studies should investigate this relationship to better understand auditory acuity during active speech and voice production.

While this exploratory study provides support for differing auditory–motor control mechanisms for speech and voice in typical speakers, study limitations should be considered when interpreting the findings. First, differences in F_1 and f_o perturbation responses may have been related to different experimental tasks or different analysis procedures used to estimate F_1 and f_o responses. Specifically, participants were not informed of gradual F_1 perturbations, sudden F_1 perturbations, or gradual f_o perturbations, although they were informed of sudden f_o perturbations. Therefore, participants’ lack of awareness of the gradual f_o perturbation and awareness of the sudden f_o perturbation may have obscured a relationship between reflexive and adaptive f_o responses. In addition, the F_1 perturbation experiments involved production of single words that required a level of linguistic processing that the f_o perturbation experiments with sustained vowels did not. Differences in analysis procedures may have also contributed to differences in adaptive and reflexive F_1 and f_o responses. For example, in the f_o adaptation experiment, responses were normalized to the baseline phase and also to the control condition to account for the natural drift in f_o seen across 108 trials; in the f_o reflex experiment, responses were normalized to the unperturbed period immediately preceding the perturbation in the same trial consistent with prior studies. In contrast, in the F_1 adaptation experiment, responses were normalized to the baseline phase only because F_1 tends to be stable across multiple productions; in the F_1 reflex experiment, responses were normalized to the average of the same word in unperturbed trials because there was no unperturbed period immediately preceding the F_1 perturbation to use for normalization. Normalizing f_o responses to the control condition altered the participants’ average response magnitudes in the perturbed condition, which may have impacted the current study findings, although the response magnitudes were similar to those found by Jones and Munhall (2000). Furthermore, this study employed a brief JND paradigm to assess auditory acuity relative to psychoacoustical studies of pitch discrimination employing hundreds of test trials (e.g., Dai & Micheyl, 2011), which may have reduced the precision of pitch acuity estimation. Additionally, the step size of 3% of the baseline for F_1 JND and 4 cents for f_o JND may not have yielded adequately precise thresholds for pitch acuity estimation. Finally, this study did not account for reflexive responses that may have been produced during the active auditory acuity experiments and may have affected JND scores. As reflexive responses to F_1 perturbation were considerably smaller than reflexive responses to f_o perturbation, estimation of auditory acuity to f_o may have been impacted more by compensatory responses to the f_o perturbations in the experiment.

Conclusions

In conclusion, this exploratory study revealed that individual speakers' responses to predictable and sudden perturbation of F_1 are related and that speakers' responses to sudden perturbation of F_1 are related to their auditory acuity to F_1 . This, combined with the finding that individual speakers' responses to predictable and sudden perturbation of f_0 were not related to each other nor to auditory acuity to f_0 , supports disparate control mechanisms for articulatory and voice parameters. Further research is warranted to investigate relationships between feedback control, feedforward control, and auditory acuity.

Acknowledgments

This research was funded by National Institute on Deafness and Other Communication Disorders Grants DC015570 (Principal Investigator [PI]: C. E. Stepp), DC016270 (PI: C. E. Stepp and Multiple PI: F. Guenther), DC015446 (PI: R. Hillman), and DC017001 (PI: R. A. Lester-Smith) and National Institute on Disability, Independent Living, and Rehabilitation Research Advanced Rehabilitation Research Grant 90AR5015 (PI: L. Cherney). We thank Talia Mittelman for her assistance with participant recruitment and data collection, Andres Llico for his technical assistance with the equipment, Yeonggwang Park for his assistance with data collection, and Frank Guenther for his thought-provoking comments on the relation of auditory perturbation responses and auditory acuity.

References

- Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLOS ONE*, 13(1), Article e0191839. <https://doi.org/10.1371/journal.pone.0191839>
- American Speech-Language-Hearing Association. (2020, October 1). *Adult Hearing Screening*. https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589942721§ion=Key_Issues
- Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America*, 132(4), 2468–2477. <https://doi.org/10.1121/1.4746984>
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Versions 5.3–6.0) [Computer software]. <http://www.praat.org>
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6), 3153–3161. <https://doi.org/10.1121/1.423073>
- Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. *Journal of Voice*, 11(2), 202–211. [https://doi.org/10.1016/S0892-1997\(97\)80079-3](https://doi.org/10.1016/S0892-1997(97)80079-3)
- Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). *A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/* [Paper presentation]. 8th International Seminar on Speech Production, Strasbourg, France.
- Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research*, 43(3), 721–736. <https://doi.org/10.1044/jslhr.4303.721>
- Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice f_0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, 121(2), 1157–1163. <https://doi.org/10.1121/1.2404624>
- Chen, X., Zhu, X., Wang, E. Q., Chen, L., Li, W., Chen, Z., & Liu, H. (2013). Sensorimotor control of vocal pitch production in Parkinson's disease. *Brain Research*, 1527, 99–107. <https://doi.org/10.1016/j.brainres.2013.06.030>
- Dai, H., & Michey, C. (2011). Psychometric functions for pure-tone frequency discrimination. *The Journal of the Acoustical Society of America*, 130(1), 263–272. <https://doi.org/10.1121/1.3598448>
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed., Vol. 127). Harper.
- Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2019). Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology*, 72(10), 2371–2379. <https://doi.org/10.1177/1747021819838939>
- García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, 38(12), 1861–1881. [https://doi.org/10.1016/S0042-6989\(97\)00340-4](https://doi.org/10.1016/S0042-6989(97)00340-4)
- Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, 72(1), 43–53. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7880914
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301. <https://doi.org/10.1016/j.bandl.2005.06.001>
- Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research*, 130(2), 133–141. <https://doi.org/10.1007/s002219900237>
- Heller Murray, E. S., & Stepp, C. E. (2020). Relationships between vocal pitch perception and production: A developmental perspective. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-60756-2>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12003512
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82. <https://doi.org/10.3389/fnhum.2011.00082>
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251. <https://doi.org/10.1121/1.1288414>
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124–132. [https://doi.org/10.1044/1058-0360\(2008\)08-0017](https://doi.org/10.1044/1058-0360(2008)08-0017)
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration

- of auditory and somatosensory feedback. *The Journal of Neuroscience*, 32(27), 9351–9358. <https://doi.org/10.1523/JNEUROSCI.0404-12.2012>
- Larson, C. R., & Robin, D. A.** (2016). Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neuroscience*, 3(1), 22–39. <https://doi.org/10.3934/Neuroscience.2016.1.22>
- Levitt, H.** (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Liu, H., & Larson, C. R.** (2007). Effects of perturbation magnitude and voice f_0 level on the pitch-shift reflex. *The Journal of the Acoustical Society of America*, 122(6), 3671–3677. <https://doi.org/10.1121/1.2800254>
- Liu, H., Wang, E. Q., Metman, L. V., & Larson, C. R.** (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PLOS ONE*, 7(3), Article e33629. <https://doi.org/10.1371/journal.pone.0033629>
- Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J. F.** (2018). Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience*, 12, 91. <https://doi.org/10.3389/fnhum.2018.00091>
- MathWorks.** (2013). MATLAB (Version 8.1) [Computer software]. The MathWorks, Inc.
- Max, L., Wallace, M. E., & Vincent, I.** (2003). *Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments* [Paper presentation]. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.
- Mollaie, F., Shiller, D. M., Baum, S. R., & Gracco, V. L.** (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, 1646, 269–277. <https://doi.org/10.1016/j.brainres.2016.06.013>
- Mollaie, F., Shiller, D. M., Baum, S. R., & Gracco, V. L.** (2019). The relationship between speech perceptual discrimination and speech production in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 62(12), 4256–4268. https://doi.org/10.1044/2019_JSLHR-S-18-0425
- Mollaie, F., Shiller, D. M., & Gracco, V. L.** (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, 28(12), 1668–1674. <https://doi.org/10.1002/mds.25588>
- Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N.** (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience*, 219, 111–119. <https://doi.org/10.1016/j.neuroscience.2012.05.042>
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B.** (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience*, 37(38), 9249–9258. <https://doi.org/10.1523/JNEUROSCI.3363-16.2017>
- Parrell, B., Ramnarayanan, V., Nagarajan, S., & Houde, J.** (2019). *The FACTS model of speech motor control: Fusing state estimation and task-based control*. bioRxiv.
- Perkell, J. S.** (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25(5), 382–407. <https://doi.org/10.1016/j.jneuroling.2010.02.011>
- Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., Vick, J., & Burton, E.** (2007). Time course of speech changes in response to unanticipated short-term changes in hearing state. *The Journal of the Acoustical Society of America*, 121(4), 2296–2311. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17471743
- Plack, C. J., & Oxenham, A. J.** (2005). The psychophysics of pitch. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 7–55). Springer.
- Purcell, D. W., & Munhall, K. G.** (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120, 966–977. <https://doi.org/10.1121/1.2217714>
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R.** (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261(1), 22–29. <https://doi.org/10.1016/j.heares.2009.12.021>
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E.** (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1–10. <https://doi.org/10.1007/s00221-004-2044-5>
- Tourville, J. A., Cai, S., & Guenther, F.** (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*, 19, 060180.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H.** (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H.** (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. <https://doi.org/10.1121/1.2773966>
- Zarate, J., & Zatorre, R. J.** (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences*, 1060(1), 404–408. <https://doi.org/10.1196/annals.1360.058>