Experimental Brain Research
https://doi.org/10.1007/s00221-022-06395-7

RESEARCH ARTICLE

Auditory and somatosensory feedback mechanisms of laryngeal and articulatory speech motor control

Hasini R. Weerathunge[1,2] · Tiffany Voon[3] · Monique Tardif[2] · Dante Cilento[2] · Cara E. Stepp[1,2,4]

Received: 11 January 2022 / Accepted: 2 June 2022

Abstract

Purpose Speech production is a complex motor task involving multiple subsystems. The relationships between these subsystems need to be comprehensively investigated to understand the underlying mechanisms of speech production. The goal of this paper is to examine the differential contributions of 1) auditory and somatosensory feedback control mechanisms, and 2) laryngeal and articulatory speech production subsystems on speech motor control at an individual speaker level using altered auditory and somatosensory feedback paradigms.

Methods Twenty young adults completed speaking tasks in which sudden and unpredictable auditory and physical perturbations were applied to the laryngeal and articulatory speech production subsystems. Auditory perturbations were applied to laryngeal or articulatory acoustic features of speech. Physical perturbations were applied to the larynx and the jaw. Pearson-product moment correlation coefficients were calculated between 1) auditory and somatosensory reflexive responses to investigate relationships between auditory and somatosensory feedback control mechanisms, and 2) laryngeal and articulatory reflexive responses as well as acuity measures to investigate the relationship between auditory-motor features of laryngeal and articulatory subsystems.

Results No statistically significant correlations were found concerning the relationships between auditory and somatosensory feedback. No statistically significant correlations were found between auditory-motor features in the laryngeal and articulatory control subsystems.

Conclusion Results suggest that the laryngeal and articulatory speech production subsystems operate with differential auditory and somatosensory feedback control mechanisms. The outcomes suggest that current models of speech motor control should consider decoupling laryngeal and articulatory domains to better model speech motor control processes.

Keywords Auditory feedback · Somatosensory feedback · Speech motor control

Introduction

Speech production is a complex motor task involving multiple subsystems. Respiratory, laryngeal, and articulatory systems coordinate to produce segmental and suprasegmental features of speech. These features of speech are monitored via auditory and somatosensory feedback. The Directions into Velocities of Articulators model (DIVA; Guenther 2016) is a physiologically validated neurocomputational model of articulatory speech motor control. DIVA consolidates speech motor control theories and empirical observations from decades of behavioral and neuroimaging research conducted on speech motor control to provide a unified platform explaining speech acquisition and production in a mathematical framework. DIVA postulates that speech production can be explained via a hybrid control system

Communicated by Bill J Yates.

✉ Hasini R. Weerathunge
hasiniw@bu.edu

1 Department of Biomedical Engineering, Boston University, Boston, MA, USA

2 Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA

3 Department of Biology, Boston, MA, USA

4 Department of Otolaryngology-Head and Neck Surgery, Boston University School of Medicine, Boston, MA, USA

Published online: 23 June 2022

combining three main control components related to speech articulation: 1) a feedforward controller that uses internally stored motor programs to produce sound, 2) an auditory feedback controller that detects errors between actual acoustic output and acoustic feature targets, and 3) a somatosensory feedback controller that detects errors between actual somatosensory (i.e., kinesthetic or proprioceptive) output and somatosensory feature targets. DIVA has been successfully used in conjunction with brain imaging and behavioral experiments to refine our understanding of the neural control of speech (Guenther 2016; Miller and Guenther 2021; Perkell 2013; Terband et al.2014 Tourville and Guenther 2011). According to the DIVA model, when there is an intent to produce a speech sound, the speech sound map selects the motor program for the intended phoneme, syllable, or word being produced. This is the desired production of the speech sound. The feedforward controller uses these internally stored motor programs to produce sound and the productions are monitored via auditory and somatosensory feedback. Disparities between sensory feedback and the desired productions of speech features are detected and used to generate real-time corrections to speech productions using sensory feedback control mechanisms (Burnett et al. 1997; Tourville et al. 2008). These corrective commands are also integrated into subsequent productions for persistent errors through feedforward control mechanisms (Tourville and Guenther 2011). Investigating the contributions of sensory feedback control mechanisms on laryngeal and articulatory systems is crucial to generate a framework of speech motor control mechanisms that can then be used as a benchmark to understand underlying variabilities in speech pathologies affecting one or more of these subsystems.

Psychophysical experiments have been used in the past to study speech motor control mechanisms. Reflexive perturbations to sensory feedback are a common paradigm used to assess the extent of an individual's reliance on the feedback modality being perturbed (Burnett et al. 1998, 1997; Burnett and Larson 2002; Chen et al. 2013; Hain et al. 2000; Lametti et al. 2012; Larson et al. 2001, 2000; Munhall et al. 2009; Parrell et al. 2017; Purcell and Munhall 2006b; Scheerer and Jones 2012). Reflexive paradigms present sudden, unpredictable perturbations to auditory or somatosensory feedback to measure the feedback-based error detection, correction, and incorporation to real-time motor production (e.g., Abbs et al. 1984; J.L. Elman 1981a, b). Feedback-based error detection has been associated with an individual's ability to differentiate minute changes in sensory stimuli. Thus, responses to reflexive perturbations are often interpreted with respect to an individual's sensitivity to differences in sensory feedback. Acuity paradigms typically identify 'just noticeable' differences (JNDs) between speech stimuli (Ghosh et al. 2010; Kewley-Port and Watson 1994; McGill and Goldberg 1968) for different auditory or somatosensory parameters. Acuity paradigms thus supplement reflexive paradigms by providing insight on an individual's sensory perception capabilities. Taken together, reflexive and acuity paradigms can be used to characterize an individual's sensory feedback control mechanisms for speech motor control. In typical speakers, prior studies report relationships between responses to auditory reflexive paradigms and acuity thresholds consistently in the articulatory subsystem but not in the laryngeal subsystem (Lester-Smith et al. 2020; Scheerer and Jones 2012), hinting at differential control mechanisms in the two domains. Feedback error correction as examined via reflexive paradigms has been found to be impaired in certain motor speech disorders including Parkinson's disease (PD) and Cerebellar Degeneration (CD) (see Weerathunge et al. (2022) for a review on atypical reflexive responses related to the laryngeal subsystem in individuals with PD and CD). Atypical reflexive responses and acuity thresholds related to the articulatory subsystem have also been found in individuals with PD (Mollaei et al. 2016, 2019). Although reflexive and acuity paradigms have been used in many studies (in speakers with typical speech as well as speakers with motor speech disorders), typically only one type of speech production subsystem (laryngeal vs. articulation) is examined in a particular group of speakers. Furthermore, due to the pragmatic difficulties in studying somatosensory responses, most studies have used only auditory feedback perturbations. To comprehensively examine the sensory feedback mechanisms at an individual level, it is crucial to assess feedback responses of multiple speech production subsystems and sensory feedback control mechanisms in the same speakers.

Auditory and somatosensory feedback control mechanisms in speech motor control

Sensory feedback-based error corrections in speech production are expected to contain weighted contributions from both auditory and somatosensory feedback control mechanisms. In the laryngeal subsystem, one study observed an increase in compensatory response magnitudes to auditory feedback perturbations of fundamental frequency (f_o) when participants' vocal folds were anesthetized to remove somatosensory feedback contributions (Larson et al. 2008). This increase was interpreted as a corrective mechanism to account for the reduced feedback from the somatosensory system caused by the vocal fold anesthesia. Thus, the authors theorized that both somatosensory and auditory feedback are used to stabilize vocal f_o shortly after perturbations. In the articulatory subsystem, Katseff et al. (2012) found that compensation in response to reflexive F_1 auditory feedback alterations was reduced when perturbation magnitude increased. In the laryngeal subsystem, Liu and Larson (2007) observed similar effects for auditory reflexive f_o perturbations. Based on these observations, Katseff et al. (2012) suggested that the auditory feedback contribution is down-weighted when the auditory feedback alteration is large and highly deviant from the unchanged somatosensory feedback. They postulated that the weighting of one sensory feedback modality increases when the reliance on the other is reduced. Alternatively, when sensory feedback has larger discrepancies compared to desired productions, the speech motor control system may lower the weighting on feedback controllers and increase reliance on the feedforward controller as the long delays in sensory feedback tend to affect the stability of motor productions (Guenther 2016). In summary, these studies suggest that the outputs of auditory and somatosensory feedback control mechanisms are not fully independent of one another or feedforward control mechanisms.

Work by Lametti et al. (2012) added substantial refinement to the understanding of the relationship between auditory and somatosensory feedback control, directly comparing responses across feedback control mechanisms in the same group of participants. Using gradual, predictable (i.e., adaptive) physical perturbations to the articulatory subsystem, they observed that the relative weight of auditory versus somatosensory feedback varied across participants, and that there was a negative relationship between somatosensory and auditory adaptive responses across participants. In a more recent study, Smith et al. (2020) also examined the relationship between somatosensory and auditory responses, but using sudden, unpredictable (i.e., reflexive) physical perturbations to the laryngeal subsystem. In their study, no statistically significant correlations were found between the magnitudes of the auditory and somatosensory reflexive responses across participants. These disparate results could be due to inherent differences between the speech production subsystem examined (i.e., laryngeal vs. articulatory) or the type of perturbation used (i.e., adaptive vs. reflexive). Thus, further studies are required to investigate the relationship between auditory and somatosensory feedback control mechanisms across both laryngeal and articulatory subsystems more comprehensively using consistent perturbation paradigms.

Laryngeal and articulatory speech motor control

There is evidence that laryngeal and articulatory subsystems may interact during speech production. The hyoid-laryngeal complex plays an essential role in vocal f_0 control, as well as in vowel articulation (Honda, 1983). The larynx tends to rise with vocal f_o, (Colton and Shearer 1971; Sapir 1978; Sapir et al. 1981) and there is a tendency for the larynx position to be higher for high vowels, as compared to low vowels (Perkell 1969). Thus, there may be inevitable interactions between the laryngeal and articulatory motor control mechanisms.

Theories of speech motor control suggest that the same underlying laryngeal and articulatory motor control mechanisms are used to produce both segmental and suprasegmental features of speech (Perkell et al. 2000). The control mechanisms are likely to include a means of accounting for interactions between segmental and suprasegmental productions. Thus, the consequences of changing the kinematics of one control mechanism are not entirely independent of the other. However, there is evidence that suggests that vocal f_0 and vowel articulation are differentially susceptible to real-time manipulations of sensory feedback (Larson et al. 2000; Purcell and Munhall 2006b). Vocal f_o tends to change rapidly with the status of auditory feedback (Lane et al. 1997; Perkell et al. 1997; Svirsky et al. 1992). However, vowel formants are less sensitive to auditory feedback changes (Cowie and Douglas-Cowie 1992; Perkell et al. 1992). It has also been shown that vocal f_0 and vowel formants can be tuned and modulated independently (H. Liu, J. Auger et al. 2010; MacDonald et al. 2011). These differential outcomes could be due to the different rates of operation of segmental and suprasegmental features of speech, as segmental features occur on a time scale that does not allow feedback-based error corrections to be incorporated whereas suprasegmental features do.

In summary, prior research suggests that auditory and somatosensory feedback control mechanisms differ across specific acoustic features (i.e., f_o and F_1), providing different relative weights to feedback control mechanisms of each feature. Comprehensively investigating these differential effects will provide insight into whether the weighting differences in auditory feedback control for different auditory features can account for the inter-subject variability in compensatory responses we observe in many altered auditory feedback studies (Burnett et al. 1998; Burnett et al. 1997; Jeffrey L Elman 1981a, b; Liu et al. 2011). In turn, this information will be useful in investigating atypical speech function of individuals with impaired speech function in laryngeal and/or articulatory motor control mechanisms (Mollaei et al. 2019). Only one prior study has investigated relationships between vocal f_o (i.e., laryngeal) and vowel F_1 (i.e., articulatory) auditory feedback control (MacDonald and Munhall 2012). The study investigated relationships between vocal f_o (i.e., laryngeal) and vowel F_1 (i.e., articulatory) adaptive response magnitudes within the same speaker. However, no statistically significant relationships were observed in the study. The authors of the study have suggested that future studies should isolate the reflexive components of the f_o and F_1 auditory feedback control to better investigate the relative weighting hypothesis. However, no prior study has investigated the relationship between auditory reflexive responses of f_o and F_1 within the same speaker. With respect to somatosensory feedback control, no studies have investigated the relationship between somatosensory reflexive responses to physical perturbations of the larynx and jaw within the same speaker. Thus, the interactions between vocal f_o and vowel F_1 need to be investigated further, incorporating both auditory and somatosensory feedback control mechanisms within the same speaker.

Another variable that should be considered when investigating compensatory responses to auditory feedback perturbations is the perceptual acuity of the speaker. Several models of speech motor control predict that speakers with finer acuity are more likely to detect and correct for errors when feedback is perturbed (Guenther 2016; Hickok 2012; Houde and Nagarajan 2011; Parrell et al. 2019). Although the relationship between perceptual acuity and compensation to auditory feedback has been explored (auditory acuity and F_1 adaptive responses: Feng et al. 2011; Lester-Smith et al. 2020; Martin et al. 2018; Villacorta et al. 2007, auditory acuity and f_o reflexive responses: Lester-Smith et al. 2020; Smith et al. 2020), the relationship between perceptual acuity of f_o and F_1 has not been researched before. In other words, are the same speakers who have higher acuity for f_o the same speakers who have higher acuity for F_1? Vowel F_1 perception is categorical in nature, whereas vocal f_o perception is more continuous in nature (Gerrits & Schouten 1998; Pisoni 1971). This would suggest that the auditory perceptual mechanisms of f_o and F_1 are not related. However, vowels are often perceived less categorically compared to consonants (Gerrits and Schouten 2004), thus the distinction between f_o and vowel F_1 may be less pronounced. Nevertheless, to the best of our knowledge, no direct investigations have been performed to compare the perceptual acuity between vocal f_o and vowel F_1. Understanding the relationship between acuity to vocal f_o and vowel F_1 is necessary to fully understand the contribution perceptual acuity has on the individual variability in responses to altered auditory feedback paradigms.

Study purpose and hypotheses

Two main research questions were investigated in the current study. The first question was whether there are relationships between responses to reflexive perturbations to auditory and somatosensory feedback. Considering conflicting results from prior studies (Lametti et al. 2012; Smith et al. 2020) and existing speech motor control models, it was hypothesized that there would be a negative relationship between the magnitudes of the responses to auditory and somatosensory feedback perturbations within the same speakers, i.e., that individuals with larger responses to auditory feedback perturbations would have smaller responses to somatosensory feedback perturbations. Thus, negative relationships were expected between responses to: H1) auditory f_o perturbations and physical perturbations to the larynx, and H2) auditory F_1 perturbation and physical perturbations to the jaw. The second question was whether there are relationships between the auditory-motor features of the laryngeal and articulatory subsystems. Based on MacDonald and Munhall (2012), it was hypothesized that there would be no statistically significant relationship between auditory-motor features of feedback control for f_o and F_1 within the same speakers. Therefore, no statistically significant correlations were expected between: H3) auditory reflexive responses of f_o and F_1, or H4) somatosensory reflexive responses of f_o and F_1. To examine hypotheses H3 and H4, we tested the alternative hypotheses that reflexive responses would be related across f_o and F_1 in auditory and somatosensory feedback control domains, respectively. H5: Based on prior research on the nature of the perception of f_o and F_1 (i.e., continuous and categorical perception respectively), no statistically significant relationships were expected between f_o and F_1 acuity within the same speaker. To examine hypothesis H5 we tested the alternative hypothesis that perceptual acuity would be related across f_o and F_1.

Materials and methods

Participants

Twenty cisgender adults aged 18–24 years (M = 21 years, SD = 2 years; 10 females, 10 males) participated in two experimental sessions. The participants were native American English speakers, with no reported history of speech, language, hearing, or neurological disorders. None of the participants had received professional training as singers or musicians. All participants completed a hearing threshold assessment [Burk and Wiley 2004; American Speech Language Hearing Association 2005)] and had thresholds ≤ 30 dB HL at octave frequencies across 250–4000 Hz. The hearing threshold assessment was administered using Radioear IDO51880 IP30 insert earphones (Radioear, New Eagle, PA) and a Grason-Stadler GSI 18 screening Audiometer (Grason-Stadler, Eden Prairie, MN). The participants provided written consent, in compliance with the Boston University Institutional Review Board.

Instrumentation and procedure

All experimental protocols were carried out in a sound-attenuating audiometric booth. The participants sat in front of a computer screen that provided visual cues. Speech was recorded using a SHURE MX153 omnidirectional condenser ear set microphone (SHURE, Niles, IL), placed 7 cm from the corner of the mouth at a 45-degree angle (Patel et al. 2018). Tasks were carried out during two 2 h sessions, generally a week apart (M = 6 days, SD = 4 days, Range = 1–17 days). The first session contained tasks to

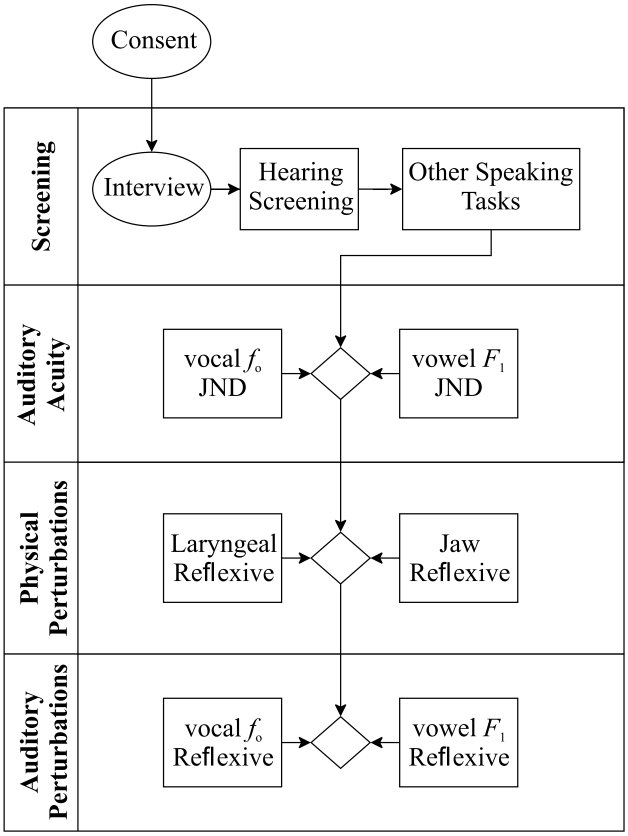

Fig. 1 Diagram illustrating the experimental protocol. There were three main sections presented in the same order for all participants. The order of presentation of sections involving f_o vs F_1 perturbations was counterbalanced across participants

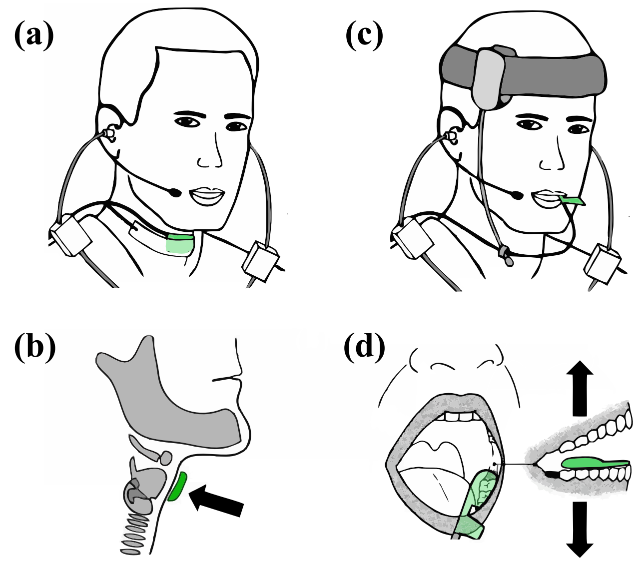

Fig. 2 Physical Perturbation Apparatus. **a** Laryngeal perturbation setup. **b** Placement of laryngeal balloon apparatus on top of the laryngeal prominence of participants. **c** Jaw perturbation setup. **d** Placement of jaw balloon apparatus between left molar teeth of participants

measure auditory acuity and speech tasks not related to the current investigation. The second session contained tasks to measure reflexive responses to auditory and physical perturbations (See Fig. 1 and Experimental procedures for more details).

All experimental paradigms were carried out via custom scripts written in MATLAB (Mathworks, Natick, MA, Version 9.4.0.813654 [R2018a]) software. The microphone signal was amplified via an RME Quadmic II microphone preamplifier (RME, Haimhausen, Germany) and was digitized either via an RME Fireface UCX sound card (for all auditory f_o and physical perturbation paradigms) or a MOTU Ultralite-mk3 Hybrid sound card (MOTU, Cambridge, MA; for auditory F_1 perturbation paradigms), both with 32-bit resolution and sampling rates of 44,100 Hz. The microphone signal was processed for f_o shifting through an Eclipse V4 Harmonizer (Eclipse, Little Ferry, NJ), which shifted all frequencies in the voice spectrum (i.e., voice harmonics and formants). The microphone signal was processed for F_1 shifting with Audapter (Cai et al. 2008), a MATLAB software package for configurable real-time manipulation of acoustic parameters of speech. All processed speech signals were amplified via a Behringer Xenyx Q802USB earphone amplifier (Behringer, Willich, Germany) and presented back to the participants over Etymotic ER-2 insert earphones (Etymotic Research, Elk Grove Village, IL) to provide near real-time auditory feedback. The total intrinsic hardware-specific processing delay was 11 ms (Heller Murray et al. 2019) for Eclipse-based f_o perturbation paradigms and 20 ms for Audapter-based F_1 perturbations paradigms. The resulting amplified signal was also digitized for offline analysis.

The physical perturbations were applied via a small, approximately tubular, inelastic balloon, constructed with heavy-duty nitrile glove material (see Fig. 2). Balloons were connected to a tube attached to a custom displacement device with a solenoid-driven air cylinder. The solenoid was computer-activated to deliver 2–4 Pounds per Square Inch (PSI) pressure and cause inflation of the balloon to a thickness of approximately 1 cm within 100 ms. The air cylinder was placed outside the sound booth to avoid exposing the participants to any auditory cues regarding the onset of physical perturbations. Microphone, earphone, pressure sensor, and computer-trigger signals were acquired by a NI USB-6212 data acquisition device (National Instruments, Austin, TX) at a sampling rate of 8000 Hz. The balloon inflation had a latency of 38 ms (SD = 5 ms). All trials of physical perturbation paradigms had speech-shaped masking noise playing back via the earphones at 75 dB SPL instead of the speech signal. The speech-shaped noise was generated using a white noise sample of 10s convolved with a spectral envelope of an /i/ vowel production. The speech-shaped noise was looped through the full duration of the physical perturbation experiments with a linear 2s fade in and fade out at the beginning and end of each paradigm. The noise was applied to ensure that the air-conducted auditory feedback of the participants' speech, along with any auditory consequences of the physical perturbation, were fully masked. Smith et al. (2020) used a similar setup and compared compensatory responses generated in response to a laryngeal physical perturbation carried out with and without masking noise. The results indicated that there was a statistically significant reduction in compensatory response magnitudes when masking noise was present. These results are congruent with the theory that having auditory feedback of the acoustic consequences of the physical perturbation causes the system to compensate for errors using both the auditory and somatosensory feedback controllers and thus increasing the total compensatory response magnitude.

Apparatus for physical perturbations

Physical perturbations of the larynx

The balloon was positioned above the cricoid cartilage and aligned with the thyroid cartilage for the larynx perturbations (Fig. 2 a, b). Inflation of the balloon pushed the larynx posteriorly and superiorly, thereby constricting the vocal folds, and reducing auditory f_o produced (See methods in Smith et al. (2020) for further details). The laryngeal balloon was attached to a flexible plastic collar, which was secured around the neck such that the balloon was placed above the cricoid cartilage and aligned with the thyroid cartilage using an adjustable flexible cord. Collar tightness was maintained such that only one finger could slip between the neck and the collar.

Physical perturbations of the jaw

The balloon was positioned between the molars on the left side of each participant's mouth for the jaw perturbations (Fig. 2c, d). Inflation of the balloon caused the jaw to be opened wider, thereby increasing the auditory F_1 produced (see methods in Golfinopoulos et al. (2011) for further detail). The jaw balloons were secured in place using headgear with a flexible arm, oriented to hold the balloon parallel to the plane of the molar teeth. Once the jaw balloon was placed, participants were guided to adjust placement to ensure that the balloons did not exert pressure on the cheek or the upper or lower molar teeth. The investigators also checked to confirm that the uninflated balloon did not obstruct speech-related movements.

Amplification, calibration, and vocalization monitoring

For all speech production tasks, the vocal intensity at the earphones was amplified 5 dB relative to the sound pressure level (SPL) at the microphone (Weerathunge et al. 2020). For all auditory acuity tasks, the vocal intensity at the earphones was maintained at 75 dB SPL. Auditory feedback amplification and calibration were carried out using the same procedure in Weerathunge et al. (2020). Participants were instructed to keep their vocalizations steady and consistent throughout the experiment and their vocal intensity and vocalization duration for each trial were monitored via a graphical display on the investigator's computer screen. During practice trials, the investigator provided feedback to the participants, by gesturing to increase or decrease volume, if their vocal intensity varied by more than 2 dB SPL from 75 dB SPL, or if the vocalization duration was shorter than 2s.

Experimental procedures

Participants completed several speaking and listening tasks for the study (Fig. 1). Speaking tasks included auditory reflexive paradigms with auditory feedback perturbations to f_o and F_1, and somatosensory reflexive paradigms with physical perturbations to the larynx and the jaw. Listening tasks included auditory acuity paradigms for f_o and F_1. The physical perturbation paradigms were carried out in session one due to the practical difficulties in attaching the physical apparatus required to produce the perturbation and considerations in place to make the rest of the study comfortable for the participants. The order of laryngeal and jaw perturbations was counterbalanced across participants to avoid order effects. The participants returned for a second session within a week (i.e., ranging from 1–7 days within first session). Carry-over effects (e.g., habituation or learning) from session one paradigms were not expected to last for more than a few trials according to previous studies examining vocal motor control in vocal f_o (Jones and Munhall 2000) and vowel F_1 (Purcell and Munhall 2006a; Tourville et al. 2008; Villacorta et al. 2007). In session two, auditory reflexive paradigms and auditory acuity paradigms were carried out. The acuity paradigms were carried out at the end of session two as they included instructions to the participants that may have alerted them to the speech features being studied (i.e., vocal f_o and vowel F_1). As acuity paradigms are passive tasks, we do not expect habituation or learning to occur at the end of the task. One exception to this order was that the vocal f_o reflexive paradigm was conducted after all acuity paradigms. This was due to the practical consideration that participants often notice the sudden vocal f_o shift and mistakenly deduce it to be an instrumentation malfunction and abruptly terminate the task. Thus, we instructed participants that sudden pitch variations may be present during experimentation. As we did not want this information to affect other reflexive paradigms of the study, we conducted the vocal f_o reflexive paradigm at the end of session two. For all trials in the speaking tasks, the participants were instructed to produce the words "bed", "head" or "Ed" for approximately 3s at a comfortable speaking voice when a visual cue of the word appeared on the screen. The visual cues contained the words written on the screen in addition to an image representative of each word. To retain participants' attention throughout the experiments, the inter-trial intervals were randomized between 1 and 2s and the words appeared in a randomized order. The participants were instructed to keep their vocalizations steady and consistent throughout the experiment. Prior to each speaking task, a practice session of nine trials was conducted to familiarize the participants with the task. Participants were instructed that their voice would be recorded for all trials. For auditory perturbation tasks, participants were advised that their voice would be played back via the earphones throughout the experiment. For physical perturbation tasks, participants were advised that masking noise would be played back via the earphones throughout the experiment.

Auditory reflexive paradigm

Auditory reflexive paradigms were conducted by applying a f_o or F_1 perturbation to auditory feedback. The respective acoustic parameter traces (i.e., f_o or F_1) of each trial for the auditory perturbations were observed to measure the response magnitude for each auditory perturbation. Each auditory reflexive paradigm contained 108 trials. Approximately half of the experimental trials were perturbed (i.e., 48 trials for f_o perturbations and 54 trials for F_1 perturbations[1]) whereas the remainder were control trials that contained no perturbation. Half of the perturbed trials (i.e., 24 trials for f_o perturbations and 27 trials for F_1 perturbations) were shifted up (+ 100 cents[2] for f_o perturbations and + 30%[3] for F_1 perturbations) whereas the remainder were shifted down (–100 cents for f_o perturbations and –30% for F_1 perturbations). The perturbation occurred with a randomized onset between 500 and 1000 ms after vocalization onset and was sustained through each perturbed trial. The 108 trials were presented in a pseudo-random order. No more than three perturbation trials occurred consecutively and adjacent perturbations were never shifted in the same direction. The first and final trials of the experiment were control trials. Prior to the beginning of the f_o reflexive paradigm, the participants were informed that there could be sudden f_o variations in their voice feedback during the experiment. This disclosure was to ensure that the participants did not terminate the experiments prematurely, believing that the equipment was malfunctioning.

Somatosensory reflexive paradigms

Somatosensory reflexive paradigms were conducted by applying a physical perturbation that affected somatosensory feedback. Physical perturbations applied to the larynx affected acoustic f_o outcomes and physical perturbations applied to the jaw affected acoustic F_1 outcomes. Thus, the respective acoustic parameter trace (i.e., f_o or F_1) of each trial for physical perturbations was observed to calculate the perturbation magnitude as well as the resulting response magnitude for each physical perturbation. Each somatosensory reflexive paradigm contained 48 trials. One-quarter of the trials (i.e., 12 trials) were perturbed and the remainder were control trials that contained no perturbation. The perturbation occurred with a randomized onset 500–1000 ms from vocalization onset and was sustained through each perturbed trial. The 48 trials were presented in a pseudo-random order. Perturbed trials were never adjacent to one another. The first and final trials of the experiment were control trials.

Acuity paradigms

Discrimination tasks designed to identify f_o and F_1 acuity were carried out by each participant. Each acuity task contained a maximum of 60 trials and a practice session of a maximum of 20 trials. Before the f_o acuity experiment, a single vocalization of the participant producing the word 'bed' for approximately one second was recorded; a one-second recording containing this full word production was used as the baseline stimulus of the experiment. Before the F_1 acuity experiment, three repeated vocalizations of the words 'bid, 'bed', 'bad' were recorded, each cued to be approximately one-second long. The production with the median F_1 formant values was selected to be used as the baseline stimulus of the experiment. Each trial contained three productions of the baseline stimuli, presented through insert earphones. Each vocal stimulus was exactly one second in duration and the interval between stimuli in each trial was 500 ms.

In each trial, the acoustic parameter (i.e., either f_o or F_1) of two of the three stimuli were perturbed equally in a randomized direction (i.e., shift up or shift down). The remaining, third stimulus was perturbed by the same amount

[1] A script error caused only 48 trials to be perturbed in all auditory f_o reflexive paradigms across all participants, instead of the expected 54 perturbed trials. The auditory F_1 reflexive paradigm contained the 54 perturbed trials as expected.

[2] 100 cents = 1 ST.

[3] The first formant value of the vowel produced (i.e. /ε/) was shifted by + 30% (i.e., towards /æ/) for F_1 shift up perturbations and -30% (i.e., towards /I/) for F_1 shift down perturbations.

in the opposite direction. The order of stimuli pairing was randomized such that either the first two stimuli or the last two stimuli of a trial contained identical perturbations; participants were informed that the different stimulus would never be the second stimulus. The participants were asked to judge whether the "first" or "last" stimulus of a trial was different from the rest. Judging two consecutive trials correctly resulted in a 'down step'—in the following trial, the difference in f_o or F_1 between the stimuli was decreased. One incorrect judgment resulted in an 'up step'— in the following trial, the difference in f_o or F_1 between the stimuli was increased (García-Pérez 1998; Levitt 1971; Macmillan and Creelman 2004). An adaptive step size was used to reduce the experimental time (Dai 1995). The first trial had a difference of 50 cents (for f_o) or 40% (for F_1) between the different stimuli. The initial step size was maintained at 10 cents (for f_o) or 5% (for F_1) until the participant provided an incorrect response. After the first incorrect response, the step size was reduced to 4 cents (for f_o) or 1.2% (for F_1) and after the participant reached a value lower than 10 cents (for f_o) or 5% (for F_1), the step size was further reduced to 1 cent (for f_o) or 0.3% (for F_1). The experiment continued until there were 10 reversals (an 'up step' followed by a 'down step' or vice versa) or 60 trials of the experiment were completed. The participants completed between 24 and 60 trials, with a mean number of 39 trials. Only two participants continued the experiment for 60 trials with one case consisting of 10 reversals and the other case consisting of six reversals.

Data analysis

Data analysis was carried out using custom analysis scripts written in MATLAB. All experimental trials were visually and audibly inspected by investigators to exclude trials that contained production errors, glottal fry, and gross formant- or f_o- tracking errors. Custom frequency bounds based on each participant's estimated vocal f_o were used to correct any f_o mistracking during secondary analysis. For all speech production trials, investigators manually marked the onset and offset of the vowel /ε / f_o traces for each trial of the auditory f_o perturbation and laryngeal perturbation experiments were calculated and extracted using the autocorrelation method via Praat software (Boersma and Weenink (2016), Versions 5–6.0.40). F_1 trajectories of each trial of auditory F_1 perturbation and jaw perturbation experiments were extracted from the data output of Audapter software (Cai et al. 2008). The f_o and F_1 trajectories of all reflexive experimental trials were extracted starting 100 ms before perturbation onset and ending 1000 ms after perturbation onset. For all perturbed trials, the 100 ms region before the perturbation onset was defined as the baseline period. For all unperturbed control trials, 400–500 ms region from vowel onset was defined as the baseline period. To exclude cross-trial variations in productions, all perturbed trials were normalized to each trial's baseline period mean f_o or F_1 magnitude[4]. To exclude natural variations in f_o and F_1 within each trial production, the baseline-normalized mean control trial trajectory (i.e., the time-series of the acoustic measure contour across time)[5] was subtracted from the baseline-normalized perturbed trials. Shift-down perturbation responses were inverted over the baseline mean to collapse perturbation response across the two directions. The resultant traces per participant were averaged to form a mean response time-trace of 600 ms, encompassing the 100 ms baseline period and a 500 ms period after the onset of perturbation. Note that for both auditory and somatosensory reflexive paradigms, the respective acoustic parameter trace (i.e., f_o or F_1) was observed to calculate response magnitudes for the respective perturbation. Physical perturbations applied to the larynx affected acoustic f_o outcomes and physical perturbations applied to the jaw affected acoustic F_1 outcomes. As the physical perturbation elicited variable perturbation magnitudes in each participant, a compensatory index, similar to Smith et al. 2020, was defined for all auditory and physical perturbation paradigms.

Auditory reflexive responses

The point in the earphone signal at which auditory perturbation was applied was manually marked as the perturbation onset. The mean response magnitude of each auditory perturbation was defined as the mean of an analysis window ranging from 120 to 240 ms from perturbation onset of each trial as per prior research (i.e., based on the latency of auditory feedback response; Lester-Smith et al. 2020; Murray and Stepp 2020; Weerathunge et al. 2020). For auditory perturbation paradigms, the compensatory index was defined to be the ratio between the mean reflexive response magnitude and the auditory perturbation magnitude (See Fig. 3). Note that the auditory perturbation magnitude was consistent for f_o (100 cents) and F_1(30%) across all participants for auditory perturbations.

Somatosensory reflexive responses

The point in time at which the neck or jaw balloon inflation initiated was manually marked as the perturbation onset.

[4] $f_{o_{normalized}}(cent) = 1200 * log_2\left(\frac{f_{o_{raw}}}{f_{o_{avgbaseline}}}\right)$, $F_{1_{normalized}}(percent) = 100 * \left(\frac{F_{1_{raw}} - F_{1_{avgbaseline}}}{F_{1_{avgbaseline}}}\right)$

[5] All control trials in reflexive paradigms were baseline normalized similar to perturbed trials. For control trials, baseline period encompassed 400 – 500 ms from vowel onset and the total control trial trajectory encompassed 400 – 1000 ms from vowel onset.

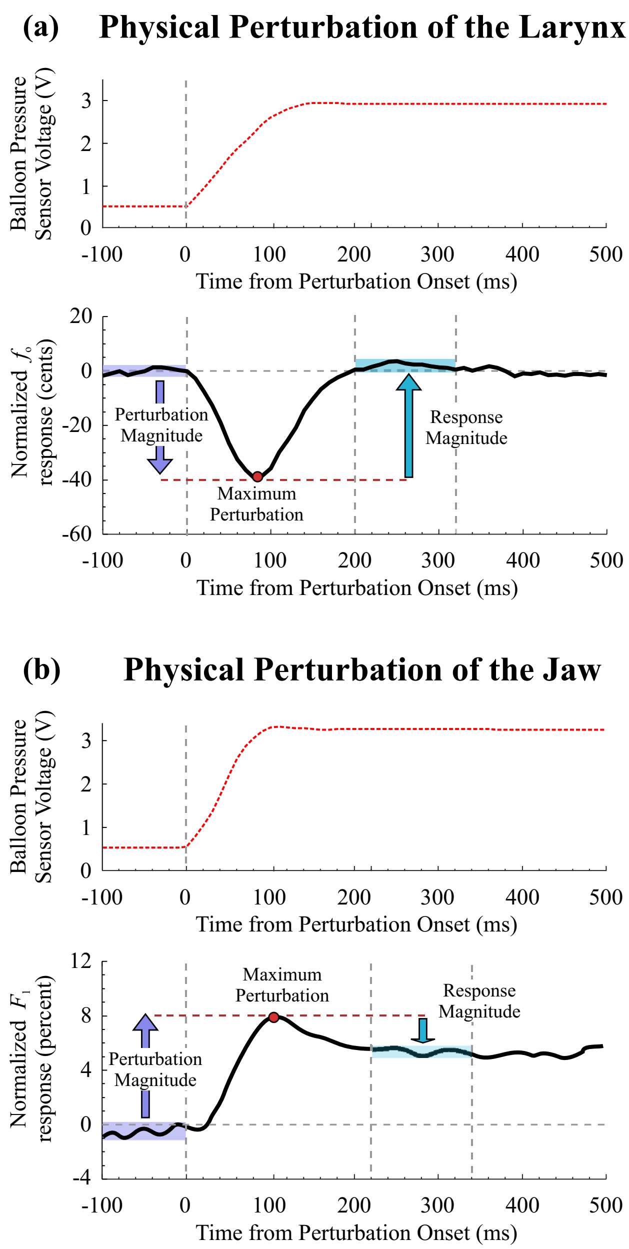

Fig. 3 Illustration of Compensatory Index calculation for physical perturbations. **a** The physical perturbation to the larynx measured via the group mean balloon pressure sensor voltage is shown in the top panel and the group mean acoustic consequences of physical perturbation of the larynx and the group mean response is shown in the bottom panel, **b** The physical perturbation to the jaw measured via the group mean balloon pressure sensor voltage is shown in the top panel and the group mean acoustic consequences of physical perturbation of the jaw and the group mean response is shown in the bottom panel. Note that the Compensatory Index (CI) for physical perturbations is calculated as the ratio between auditory response magnitude and the perturbation magnitude (i.e., acoustic consequence of the physical perturbation)

The mean response magnitude of each physical perturbation was defined as the difference between the mean of an analysis window ranging from 120 ms to 240 ms from perturbation onset and the peak perturbation magnitude in the response time-trace. This time window was selected as the balloon inflation had a latency of 38 ms (SD = 5 ms) from perturbation onset (i.e., latency to obtain peak physical perturbation), and the acoustic consequence latency of somatosensory feedback has been observed to be 55–65 ms (Sapir et al. 2000) and 70–80 ms (Loucks et al. 2005) according to prior research. Thus, we expected somatosensory feedback-related modifications would be present in the acoustic signal at around 120 ms after perturbation onset (also confirmed via visual inspection). The peak perturbation magnitude was interpreted as the point at which the physical perturbation caused the largest acoustic effect relative to the baseline (Loucks et al. 2005; Sapir et al. 2000; Smith et al. 2020). The compensatory index for physical perturbation paradigms was the ratio between the mean reflexive response magnitude and the peak perturbation magnitude (See Fig. 4)[6]. Perturbation magnitudes smaller than the baseline variability (i.e., the standard deviation of baseline mean) were considered to have no perturbation effect and were thus removed from further analysis.

Acuity

The Just Noticeable Difference (JND) threshold for the f_o and F_1 acuity tasks was defined as the mean f_o value (in cents) or the mean F_1 value (in percentage) of the final six reversals out of the ten reversals per participant. Except for one participant who had six reversals for their F_1 acuity paradigm and ten reversals for their f_o acuity paradigm, all other participants had 10 reversals in both f_o and F_1 acuity paradigms. In this specific case, the final six reversals' mean value was still considered as the JND threshold.

Statistical analysis

All participant responses were used for statistical analysis regardless of the nature of the response (i.e., opposing or following the direction of the perturbation) because all response types cumulatively characterize speech motor control behavior of the participants (Behroozmand et al. 2012; Franken et al. 2018; Li et al. 2013; Patel et al. 2014). All data were tested for normality and homogeneity using Anderson–Darling tests. As data from one of the six measures was not normal, nonparametric tests were conducted for statistical analysis of all data. To investigate the relationship between

[6] $Compesatory\ Index\ (CI) = \left(\frac{Mean\ reflexive\ response\ magnitude}{Maximum\ perturbation}\right)$

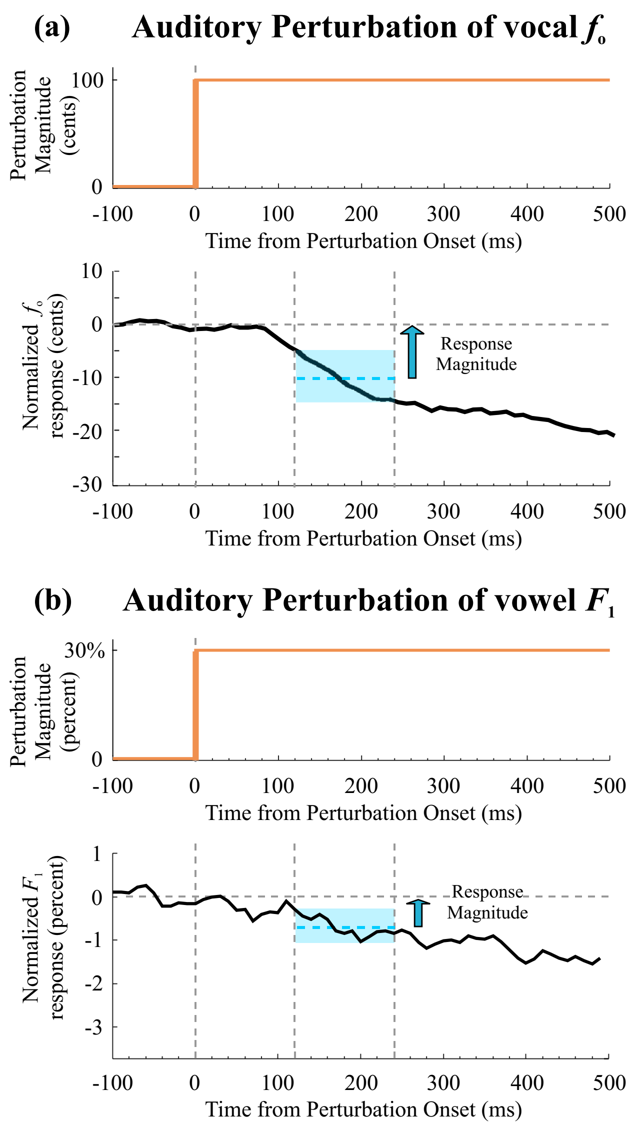

Fig. 4 Group mean responses for auditory perturbations. **a** The auditory f_o perturbation is shown in the top panel and the group mean response for auditory f_o perturbation is shown in the bottom panel, and **b** the auditory F_1 perturbation is shown in the top panel and the group mean response for the auditory F_1 perturbation is shown in the bottom panel. Note that the Compensatory Index (CI) for auditory perturbations is calculated as the ratio between auditory response magnitude and auditory perturbation magnitude

auditory and somatosensory feedback control mechanisms, within each speaker, Spearmen's rank correlation coefficients were calculated 1) between the compensatory indices of auditory and somatosensory f_o reflexive responses, and 2) between the compensatory indices of auditory and somatosensory F_1 reflexive responses. To investigate the relationship between auditory-motor features of laryngeal and articulatory subsystems, within each speaker, Spearmen's rank correlation coefficients were calculated 1) between the compensatory indices of auditory reflexive responses for f_o and F_1 auditory perturbations, 2) between the compensatory indices of somatosensory reflexive responses physical perturbations of the larynx and the jaw, and 3) between auditory f_o and auditory F_1 acuity.

Mann–Whitney tests were performed on the subset of participants ($n = 17$ out of 20) for which both auditory and somatosensory reflexive responses were present for laryngeal perturbations to compare auditory and somatosensory reflexive response compensatory indices in auditory f_o and laryngeal physical perturbations. Similarly, Mann–Whitney tests were performed on the subset of participants ($n = 18$ out of 20) for which both auditory and somatosensory reflexive responses were present for articulatory perturbations to compare auditory and somatosensory reflexive response compensatory indices in auditory F_1 and jaw physical perturbations. Although these tests were not related to the main study questions or related to an explicit hypothesis a priori, given the small number of extant studies that measure both auditory and somatosensory responses, we wanted to document the difference between the two types of responses.

A significance level of $p < 0.05$ was applied to analyses. All statistical analyses were conducted in Minitab (Ryan et al. 2014; Version 2019).

Results

On average, participants produced compensatory responses (i.e., in the direction opposite to the perturbation) to auditory and physical perturbations of f_o and F_1 (see Fig. 4). The group median reflexive response compensation index for auditory f_o perturbations was 0.075 (IQR = 0.103). One participant produced a following f_o response (i.e., responses with a compensatory index less than zero), with a compensation index of − 0.003. All other participants produced compensatory f_o responses, with compensation indices ranging between 0.015 and 0.238. The group median reflexive response compensation index for auditory F_1 perturbations was 0.015 (IQR = 0.043). Six participants produced following F_1 responses, with compensation indices ranging between − 0.016 and 0. All other participants produced compensatory F_1 responses, with compensation indices ranging between 0.006 and 0.077.

The group median reflexive f_o response compensation index for physical perturbations of the larynx was 0.788 (IQR = 0.311). Three participants displayed no measurable perturbation magnitude for vocal f_o (i.e., an indication that the physical perturbation was not successful) and thus had to be removed from the analysis. All remaining participants produced compensatory f_o response compensation indices, with average values between 0.022 and 0.887. The group median reflexive F_1 response compensation index for physical perturbations of the jaw was 0.421 (IQR = 0.562). Two participants were removed from the analysis, as there was no measurable perturbation magnitude for F_1 (i.e., an indication that the physical perturbation was not successful). All other participants produced compensatory F_1 response compensation indices, with values between 0.007 and 1.401.

Mann–Whitney tests performed to compare auditory and somatosensory reflexive response compensatory indices in auditory f_o and laryngeal physical perturbations showed that compensatory indices for auditory f_o responses (median = –0.072) were significantly lower than compensatory indices for laryngeal physical perturbation responses (median = 0.788); W(17) = 153, $p < 0.001$. Similarly, Mann–Whitney tests performed to compare auditory and somatosensory reflexive response compensatory indices in auditory F_1 and jaw physical perturbations showed that compensatory indices for auditory F_1 responses (median = 0.015) were significantly lower than compensatory indices for jaw physical perturbation responses (median = 0.421); W(18) = 189, $p < 0.001$. These results indicated that the partial compensations for the physical perturbations were statistically significantly larger in magnitude compared to auditory perturbations. See Fig. 5 for a visual comparison between auditory and somatosensory feedback response compensatory indices. The median f_o JND threshold was 18.6 cents (IQR = 16.3 cents, Max = 46.0 cents, Min = 6.2 cents) and the median F_1 JND threshold was 3.70% (IQR = 2.52%, Max = 11.00%, Min = 0.92%).

None of the correlations concerning the relationships between auditory and somatosensory feedback were statistically significant. The median reflexive f_o response compensation indices for auditory f_o perturbations and physical perturbations of the larynx were not statistically significantly correlated (r(17) = 0.324, p = 0.205; see Fig. 6(a)). The median reflexive F_1 response compensatory indices for auditory F_1 perturbations and physical perturbations of the jaw were also not statistically significantly correlated (r(18) = –0.216, p = 0.390; see Fig. 6(b)). None of the calculated correlations concerning the relationships between auditory-motor features in the laryngeal and articulatory subsystems were statistically significant. The median auditory reflexive response compensation indices for f_o and F_1 were not correlated (r(20) = –0.117, p = 0.622; see Fig. 6(c)). The median somatosensory reflexive response compensation indices for physical perturbations of the larynx and the jaw articulator were also not statistically significantly correlated (r(16) = 0.459, p = 0.074; see Fig. 6(d)). The median JND thresholds between f_o and F_1 were also not correlated (r(20) = 0.177, p = 0.455; see Fig. 6(e)).

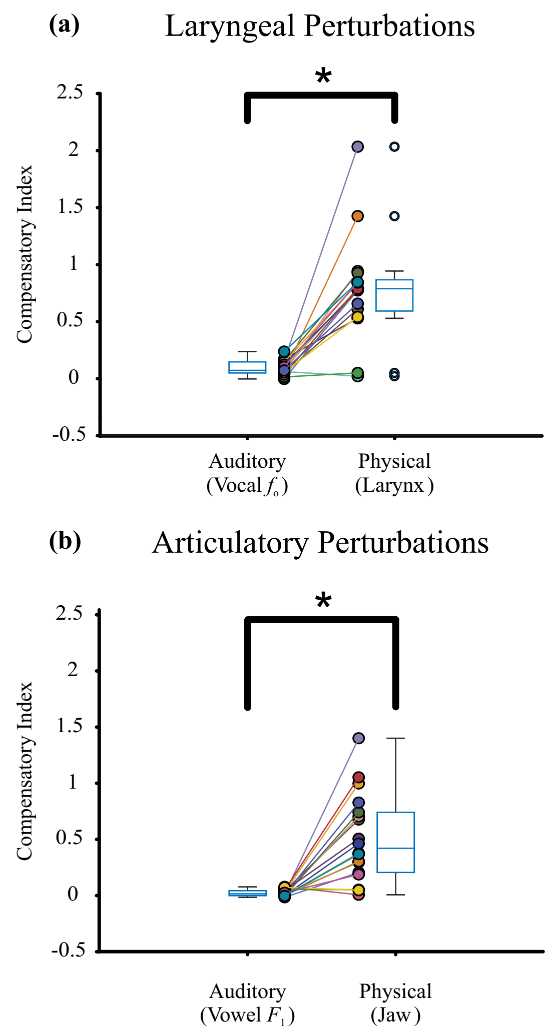

Fig. 5 Comparison of auditory and somatosensory feedback response compensatory indices. Box plots of group response compensatory indices and individual participant response compensatory indices for **a** vocal f_o vs. laryngeal perturbations, and **b** vowel F_1 vs. jaw perturbations

Discussion

In this study, we investigated two main research questions. First, whether there are relationships between responses to reflexive perturbations to auditory and somatosensory feedback. Second, whether there are relationships between the auditory-motor features of the laryngeal and articulatory subsystems. We hypothesized that there would be a negative relationship between auditory and somatosensory feedback control mechanisms, and that there would be no statistically significant relationships between the auditory-motor features of the laryngeal and articulatory subsystems.

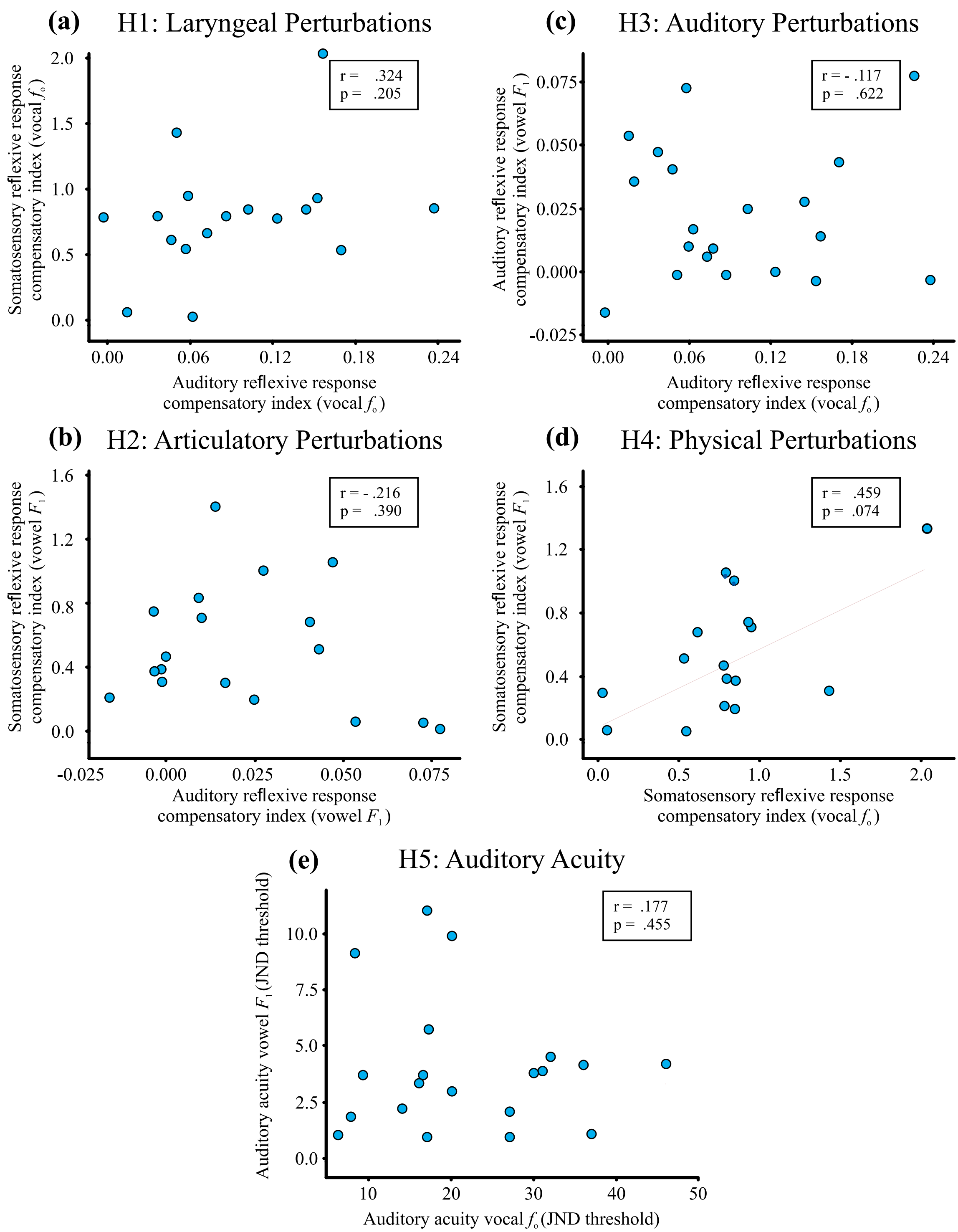

Fig. 6 Scatterplots related to Spearman's rank correlation coefficient calculations. **a** Laryngeal reflexive response compensatory indices of physical vs auditory perturbations, **b** Articulatory reflexive response compensatory indices of physical vs auditory perturbations, **c** Reflexive response compensatory indices of auditory perturbations to vocal f_o vs vowel F_1, **d** Reflexive response compensatory indices of physical perturbations to the larynx vs the jaw, **e** Just Noticeable Difference (JND) thresholds between f_o and F_1

Relationships between auditory and somatosensory feedback

In the current study, no statistically significant correlations were observed between within-speaker responses to reflexive perturbations to auditory and somatosensory feedback. Thus, our hypothesis that there would be a negative relationship between auditory and somatosensory feedback was not supported. However, our results in the laryngeal domain are congruent with the only other study that examined correlations between reflexive responses to auditory and physical perturbations of the larynx (Smith et al. 2020). The current study is the first to examine correlations between auditory and somatosensory reflexive responses to auditory and physical perturbations of the jaw. Two studies have carried out adaptive paradigms of auditory and physical perturbations of the jaw (Feng et al. 2011; Lametti et al. 2012). Lametti et al. (2012) found a statistically significant negative correlation between the adaptive responses to auditory and physical perturbations. However, the physical perturbation carried out in Lametti et al., had no acoustic consequences (i.e., the jaw was moved in a lateral direction, although the vertical movement of the jaw is instrumental for vowel F_1 changes). These specific differences in paradigm design could be possible reasons for the incongruent results we observed in the current study. Although Feng et al. (2011) conducted physical perturbations in the jaw in a vertical direction along with auditory F_1 perturbations, no correlation analysis was carried out.

As per neurocomputational models of speech such as DIVA, the contributions of auditory and somatosensory feedback control mechanisms to consequent productions are considered to be a linear-weighted summation. Thus, the relationship between the two feedback control mechanisms is expected to have a negative correlation. This is directly applicable to and can be measured via, responses to adaptive paradigms in which the contributions of feedback control mechanisms to subsequent productions are considered. This is supported by results from correlations between responses to auditory and physical perturbations (Lametti et al. 2012). However, Feng et al. (2011) speculated that the weighting of each subsystem may be dynamically adjusted across the adaptive paradigms, as the results of their study could not be fully accounted for by the linear weighted sum narrative. Moreover, the current study examined reflexive paradigms in auditory and somatosensory feedback subsystems. Thus, the responses to the paradigms are applicable to the contributions of each subsystem to real-time speech production output. From the current results, the relationship between auditory and somatosensory feedback subsystems does not seem to be linear.

During physical perturbations to the jaw, auditory feedback was masked. Thus, there was no competition between the two feedback control mechanisms' corrective commands. However, during auditory perturbation paradigms, the corrective commands generated by the auditory feedback were in competition with the error control commands produced by the somatosensory feedback, which would have sensed unaltered signals. This is likely the reason that the response compensatory indices for the physical perturbations were statistically significantly larger compared to auditory perturbations.

Several factors could contribute to the differential contribution of auditory vs. somatosensory feedback control mechanisms to speech production. Auditory feedback is expected to be more delayed compared to somatosensory feedback as auditory feedback signals are transmitted via the auditory cortex whereas somatosensory feedback signals are transmitted via the cranial nerves in the brainstem (Loucks et al. 2005). For instance, prior studies have observed shorter latencies in somatosensory feedback responses (65–75 ms; Sapir et al. 2000) compared to auditory feedback responses (100–150 ms; Hain et al. 2000). This could be a reason why auditory perturbations of suprasegmental features such as vocal f_o are more susceptible to auditory feedback perturbations compared to segmental features such as vowel F_1. For example, studies carried out on cochlear implant users who became profoundly deaf as adults show that average values of vocal f_o change relatively rapidly with the change in hearing status, whereas vowel formants remain relatively unaffected (Cowie and Douglas-Cowie 1992; Perkell et al. 1992; Svirsky et al. 1992). Another factor for differential contribution could be the amount of task-relevant information provided by each sensory modality. For instance, closed vowels contain better-specified somatosensory information compared to open vowels, and thus we can posit that closed-vowel productions rely more on somatosensory feedback contributions. This is confirmed by auditory perturbation studies performed in vowel formants in which closed vowels elicited fewer compensatory responses compared to open vowels (closed vowel: Mitsuya et al. 2015; open vowels: Purcell and Munhall 2008; Reilly and Dougherty 2013).

Apart from the above factors, research also suggests that there is a sensory preference for auditory or somatosensory feedback that varies across individuals. Lametti et al. (2012) observed a sensory preference that varied across participants in speech production when simultaneous alteration of auditory and somatosensory feedback was applied to the jaw articulator (i.e., individuals who had larger responses to one sensory feedback perturbation had smaller responses to the other sensory feedback modality perturbation). However, Lametti et al. (2012) used an adaptive paradigm and the reflexive paradigms in the current study did not display this inverse relationship between auditory and somatosensory feedback. Currently, there is limited information on which feedback modality plays a dominant role in reflexive paradigms carried out within individuals. Thus, in future studies, it is important to look at simultaneous reflexive perturbations of auditory feedback and somatosensory feedback in congruent and incongruent directions so that competitive as well as cooperative compensatory strategies of the dual systems are investigated.

Relationships between auditory-motor features in laryngeal and articulatory subsystems

There were no statistically significant correlations observed between auditory-motor features in the laryngeal and articulatory subsystems as hypothesized. There could be several reasons why the segmental feature of vocal F_1 is controlled in a manner different from that of the suprasegmental feature of vocal f_o. For instance, the control mechanisms may be different for segmental and for suprasegmental features due to the different rates the speech production subsystems operate in and the manner of production of these specific acoustic features. On one hand, the rate of error correction needs to happen faster in vowel F_1 production compared to vocal f_o production, due to the segmental nature of vowel productions. In this regard, vowel formant corrections may not incorporate auditory feedback to as large an extent and may even rely mainly on feedforward motor plans and somatosensory feedback. On the other hand, open vowel productions may contain minor contributions from somatosensory feedback due to the open configuration of the vocal tract. The speaker may rely more on auditory feedback in this case to monitor real-time productions. Vocal f_o control may require a major contribution of somatosensory feedback at the focal folds to control the tension of intrinsic muscles to vary vocal f_o productions. These subtle differences in the acoustic features predict that the way each feature is controlled by the speech motor control system may vary, and that may be reflected in the correlations calculated between the auditory-motor features in production and perception paradigms.

The results of the current study highlight the importance of investigating segmental and suprasegmental features separately with respect to their motor control mechanisms. Current models of speech motor control have been traditionally developed to focus on the articulatory subsystem and segmental features such as vowel formants (Guenther 2016; Houde and Nagarajan 2011). However, current study results are an indicator that laryngeal and articulatory subsystems may be inherently different in their motor control mechanisms. Their interactions seem to be complex and need to be investigated in a more comprehensive manner in future studies. Thus, each subsystem should be clearly and separately represented in modeling frameworks. Recently there have been efforts to incorporate suprasegmental features as controlled variables in these models (Houde and Chang 2015; Weerathunge et al. in review).

Limitations and Future Directions

There are several limitations to the current study. The physical perturbations administered across participants were of varying magnitudes. The dimensions of each individual's articulatory and laryngeal structures factored into the level of perturbation they experienced, even though the balloon apparatus used and level of pressure exerted were maintained across participants. We normalized this variability by considering a compensatory index in the statistical analysis of the auditory responses to the physical perturbations. A further limitation in the study was the mismatched number of trials used for auditory perturbation experiments vs. physical perturbation experiments. The reduced number of trials in somatosensory perturbation experiment and the restrictions on counterbalancing auditory and physical perturbation paradigms were due to the limited amount of time the custom displacement device could be operated continuously without the solenoids heating, as well as the practical difficulties in having a physical apparatus attached to the neck or placed inside the jaw of participants for longer periods of times.

Studying each sensory feedback control mechanism in isolation is ideal to understand the contributions of each system. Masking noise was applied to block auditory feedback during physical perturbation paradigms with this objective. As applying a direct somatosensory feedback perturbation to an articulator is infeasible, we carried out physical perturbations that elicit changes in somatosensory feedback by changing the position of a specific articulator (i.e., jaw or larynx). However, as the physical perturbations also elicit acoustic consequences, and thereby changes auditory feedback, the participants may compensate for auditory feedback errors in addition to somatosensory feedback errors. Note that both compensatory responses will be in the same direction and thus additive. Thus, to isolate and investigate the somatosensory feedback controller, the acoustic consequences of the physical perturbations were blocked by applying masking noise as auditory feedback to the participants. In contrast, when auditory perturbations are presented, somatosensory feedback remains unaffected. Prior research carried out using both auditory and somatosensory reflexive paradigms follow similar standardized protocols with the above difference in experimental design between auditory and somatosensory perturbations (Feng et al. 2011; Golfinopoulos et al. 2011; Nasir and Ostry 2006; Smith et al. 2020). Although the magnitude of the auditory perturbation is identical across participants, the effect of the physical perturbations on the larynx and the jaw are variable across participants due to implementation methodology and anatomical differences. We have partially resolved this variability by defining a compensatory index (i.e., the ratio between response magnitude and perturbation magnitude) to quantify and normalize the responses to each type of perturbation. Thus, the physical perturbations in the current study contain responses from the somatosensory feedback controller whereas auditory perturbations contain responses from auditory feedback controllers, and variabilities across auditory and physical perturbation paradigms are mitigated by normalizing response magnitudes by perturbation magnitudes. However, given that there is not necessarily a one-to-one correspondence between motor kinematics and generated acoustic consequences, the extent to which participants compensated for the physical perturbation via motor kinematics remains unexplored.

One argument that can be made to explain the absence of a correlation between reflexive responses to auditory and somatosensory feedback perturbations in the articulatory subsystem is the possible tradeoff of tongue versus jaw in response to the physical perturbation of the jaw. However, Feng et al. (2011) conducted mechanical perturbations of jaw in a similar manner to the current study and measured both jaw and tongue position kinematics throughout the experiment. They observed that when a physical perturbation was applied to the jaw (upwards or downwards) in the absence of auditory perturbations, the jaw perturbation was compensated by the movement of the jaw in the opposing direction while the tongue position remained unchanged. Although we did not specifically measure kinematics of the jaw and tongue during perturbations of the current study, the same mechanisms can be speculated to have occurred due to the similarity in physical perturbation. We further speculate that minimal movement of the tongue would be observed for physical perturbations of the jaw as the auditory consequences of the perturbations are masked from the participants' auditory feedback. Nevertheless, the tradeoff in compensatory articulator for the physical perturbation is likely to have had minimal effects on the measurement of the response to the physical perturbation in the current study as the acoustic consequences were used to quantify response magnitudes.

Based on the results of the current study, it is apparent that the relationships between auditory and somatosensory feedback control mechanisms and laryngeal and articulatory subsystems are complex in nature and need to be comprehensively investigated in the future. Prior research including auditory feedback perturbation studies (Abur et al. 2021; Mollaei et al. 2016) provide evidence that the pathology of PD affects laryngeal vs. articulatory speech subsystems differentially. Prior studies using acoustic measures have observed that voice symptoms emerge at the early stages of PD and are not correlated with disease severity whereas articulatory symptoms increase with disease progression (Harel et al. 2004; Midi et al. 2008; Miller et al. 2011; Skodda et al. 2012). The somatosensory feedback subsystem has yet to be comprehensively studied in PD. Future studies replicating the current study methodology to understand the differential effects of auditory and somatosensory feedback subsystems on laryngeal vs. articulatory speech subsystems in neurological disorders such as PD can provide insight into the underlying pathophysiology of those disorders which could lead to robust clinical interventions focused on each subsystem. In terms of models of speech motor control model, separate models of articulatory and laryngeal motor control have not been enacted. For example, in the DIVA model (Guenther 1994, 2016), articulatory motor control is clearly defined and the same mechanisms are extended to laryngeal motor control without consideration for possible biomechanical and/or neural changes in the way laryngeal motor control may be handled by the speech motor control system. The state feedback control (SFC) model, another well-established model, also defines speech motor control considering vocal tract articulators Houde and Nagarajan 2011. More recently, considerations have been made in defining vocal motor control separately for state feedback control model architecture using recent neuroimaging findings on vocalizations Houde and Chang 2015. Based on the current study outcomes, decoupling laryngeal and articulatory domains when modeling speech motor control processes seems to be a critical step in understanding the underlying variations of control mechanisms in each subsystem.

Somatosensory acuity is one characteristic not investigated in the current study. Somatosensation can be divided into several subcategories that include vibrational, tactile (fine and coarse), and proprioceptive somatosensation (Gritsyk et al. 2021). In the context of vocal production in the current study, laryngeal somatosensory feedback may rely heavily on tactile somatosensory feedback of the mechanoreceptors of the vocal folds (Hammer and Barlow 2010; Romo et al. 2002). On the other hand, for open vowel production, articulatory somatosensory feedback may rely more on proprioceptive feedback from the closing and opening movement of the jaw. Although there are existing somatosensory acuity paradigms that may be applicable for the physical perturbation carried out in the current study for the jaw articulator (Daliri et al. 2013; Gritsyk et al. 2021), the exact somatosensory acuity paradigm relevant to the physical perturbations carried out in the current study for the larynx (i.e., pressure applied on the laryngeal prominence) has not been tested to investigate acuity (Loucks et al. 2005). Currently, the only known laryngeal somatosensory acuity paradigm measures laryngeal mechanosensory detection, in which an air burst stimulus is applied to the laryngeal mucosa and the participant is required to acknowledge feeling the stimulus. (Hammer 2009). However, this acuity paradigm is a relatively invasive endoscopic procedure that is typically not carried out simultaneously with voicing. As we were interested in observing responses to a task-related perturbation of the larynx, a similar type of laryngeal perturbation was not pursued in the current study.

As the differences between laryngeal and articulatory speech subsystems may be driven by different rates of operation for segmental and suprasegmental features of speech, there could be more similarities between laryngeal and articulatory speech subsystems in a language that has lexical tone (i.e., where f_o control may be comparably more task-specific). We predict that there would be more restricted control on vocal f_o (i.e., similar to F_1) and that the reliability on auditory vocal f_o feedback would be higher for speakers of tonal languages. The reliability on f_o auditory feedback is evidenced by research that shows increases in the percentage of compensatory responses as the perturbation magnitude increases in total language speakers, which is in contrast to the reduction in the percentage of compensatory responses often observed in native English speakers (Hanjun Liu et al. 2010a, b, c). Tonal language speakers with PD tend to produce even larger responses to reflexive vocal f_o perturbations compared to non-tonal language speakers with PD (Chen et al. 2007; H. Liu, E. Q. Wang, et al., 2010), which suggests more restricted and task-relevant control of vocal f_o. These findings imply that the relationship between laryngeal and articulatory speech subsystems may be modulated by the language experience of an individual.

Conclusion

The current study measured auditory-motor, somatosensory-motor and acuity data in a population of young adults with typical speech to characterize the auditory and somatosensory feedback control mechanisms and the laryngeal and articulatory subsystems in individuals. The results of the study failed to confirm the hypothesis that there is a relationship between the auditory and somatosensory feedback control mechanisms of individual speakers. However, the results confirmed the hypothesis that laryngeal and articulatory speech production subsystems operate with differential auditory and somatosensory feedback control mechanisms. The study outcomes suggest that laryngeal and articulatory subsystems have differential auditory and somatosensory feedback control mechanisms. Further research is warranted to study these differential effects and as a first step, we suggest that current models of speech motor control should consider decoupling laryngeal and articulatory domains to better model speech motor control processes.

Acknowledgements We acknowledge Frank Guenther, Andrés F. Salazar-Gómez, and Manuel Díaz Cádiz for design and technical support related to the custom displacement device used to generate physical perturbations. We acknowledge Defne Abur for sharing expertise on formant perturbations, Dante Smith and Elaine Kearney for providing support in setting up laryngeal and jaw perturbation paradigms respectively, and Ashley Mcfarlane for assistance with data analysis. We would also like to thank Joseph Perkell for their critical insights on somatosensory perturbations.

Author contributions Conceptualization: CES Funding acquisition: CES Resources: CES Supervision: CES Data curation: HRW, DC Formal analysis: HRW, TV, MT Investigation: HRW, MT, DC. Methodology: HRW, CES Software: HRW Visualization: HRW, CES. Writing – original draft: HRW, TV, MT, DC, CES Writing – review & editing: HRW, CES.

Funding This research was supported by grant R01 DC016270 (CES & FHG) and by a Graduate Fellow Award from the Rafik B. Hariri Institute for Computing and Computational Science and Engineering (HRW).

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

Abbs JH, Gracco VL, Cole KJ (1984) Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. J Mot Behav 16(2):195–231. https://doi.org/10.1080/00222895.1984.10735318

Abur D, Subaciute A, Daliri A, Lester-Smith RA, Lupiani AA, Cilento D, Enos NM, Weerathunge HR, Tardif MC, Stepp CE (2021) Feedback and feedforward auditory-motor processes for voice and articulation in parkinson's disease. J Speech Lang Hear Res 64(12):4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153

American Speech Language Hearing Association. (2005). Guidelines for manual pure-tone threshold audiometry. https://www.asha.org/policy/gl2005-00014/

Behroozmand R, Korzyukov O, Sattler L, Larson CR (2012) Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control. J Acoust Soc Am 132(4):2468–2477. https://doi.org/10.1121/1.4746984

Boersma, P., & Weenink, D. (2016). *Praat:* Doing phonetics by computer (Version 6.0.14). https://www.fon.hum.uva.nl/praat/

Burk MH, Wiley TL (2004) Continuous versus pulsed tones in audiometry. Am J Audiol 13(1):54–61. https://doi.org/10.1044/1059-0889(2004/008)

Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am 103(6):3153–3161. https://doi.org/10.1121/1.423073

Burnett TA, Larson CR (2002) Early pitch-shift response is active in both steady and dynamic voice pitch control. J Acoust Soc Ame 112(3):1058–1063

Burnett TA, Senner JE, Larson CR (1997) Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. J Voice 11(2):202–211. 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 trajectories and results from perturbations of the Mandarin triphthong/iau. Proceedings of the 8th ISSP 65–68.

Chen SH, Liu H, Xu Y, Larson CR (2007) Voice F0 responses to pitch-shifted voice feedback during English speech. J Acoust Soc Am 121(2):1157–1163. https://doi.org/10.1121/1.2404624

Chen X, Zhu X, Wang EQ, Chen L, Li W, Chen Z, Liu H (2013) Sensorimotor control of vocal pitch production in Parkinson's disease. Brain Res 1527:99–107. https://doi.org/10.1016/j.brainres.2013.06.030

Colton, R. H., & Shearer, W. M. (1971). Hyoid position as a function of fundamental frequency in the modal and falsetto registers. State University of New York Upstate Medical Center.

Cowie, R., & Douglas-Cowie, E. (1992). Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences (Mouton de Gruyter, New York). Google Scholar, 304.

Dai H (1995) On measuring psychometric functions: a comparison of the constant-stimulus and adaptive up-down methods. J Acoust Soc Am 98(6):3135–3139. https://doi.org/10.1121/1.413802

Daliri A, Prokopenko RA, Max L (2013) Afferent and efferent aspects of mandibular sensorimotor control in adults who stutter. J Speech Lang Hear Res 56(6):1774–1788. https://doi.org/10.1044/1092-4388(2013/12-0134)

Elman JL (1981a) Effects of frequency-shifted feedback on the pitch of vocal productions. J Acoust Soc Am 70(1):45–50. https://doi.org/10.1121/1.386580

Elman JL (1981b) Effects of frequency-shifted feedback on the pitch of vocal productions. J Acoust Soc Am 70(1):45–50

Feng Y, Gracco VL, Max L (2011) Integration of auditory and somatosensory error signals in the neural control of speech movements. J Neurophysiol 106(2):667–679. https://doi.org/10.1152/jn.00638.2010

Franken MK, Acheson DJ, McQueen JM, Hagoort P, Eisner F (2018) Opposing and following responses in sensorimotor speech control: why responses go both ways. Psychon Bull Rev 25(4):1458–1467. https://doi.org/10.3758/s13423-018-1494-x

M A GP (1998) Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. Vision Res 38(12):1861–1881. https://doi.org/10.1016/s0042-6989(97)00340-4

Gerrits, E., & Schouten, B. (1998). Categorical perception of vowels. Fifth International Conference on Spoken Language Processing

Gerrits E, Schouten ME (2004) Categorical perception depends on the discrimination task. Percept Psychophys 66(3):363–376

Ghosh SS, Matthies ML, Maas E, Hanson A, Tiede M, Menard L, Guenther FH, Lane H, Perkell JS (2010) An investigation of the relation between sibilant production and somatosensory and auditory acuity. J Acoust Soc Am 128(5):3079–3087. https://doi.org/10.1121/1.3493430

Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH (2011) fMRI investigation of unexpected somatosensory feedback perturbation during speech. Neuroimage 55(3):1324–1338. https://doi.org/10.1016/j.neuroimage.2010.12.065

Gritsyk O, Kabakoff H, Li JJ, Ayala S, Shiller DM, McAllister T (2021) Toward an index of oral somatosensory acuity: comparison of three measures in adults. Perspect ASHA Spec Interes Groups 6(2):500–512. https://doi.org/10.1044/2021_persp-20-00218

Guenther FH (1994) A neural network model of speech acquisition and motor equivalent speech production. Biol Cybern 72(1):43–53

Guenther FH (2016) Neural control of speech. Mit Press. https://doi.org/10.7551/mitpress/10471.001.0001

Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK (2000) Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Exp Brain Res 130(2):133–141. https://doi.org/10.1007/s002219900237

Hammer MJ (2009) Design of a new somatosensory stimulus delivery device for measuring laryngeal mechanosensory detection thresholds in humans. IEEE Trans Biomed Eng 56(4):1154–1159. https://doi.org/10.1109/TBME.2008.2007968

Hammer MJ, Barlow SM (2010) Laryngeal somatosensory deficits in Parkinson's disease: implications for speech respiratory and phonatory control. Exp Brain Res 201(3):401–409. https://doi.org/10.1007/s00221-009-2048-2

Harel B, Cannizzaro M, Snyder PJ (2004) Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn 56(1):24–29

Heller Murray ES, Lupiani AA, Kolin KR, Segina RK, Stepp CE (2019) Pitch shifting with the commercially available eventide eclipse: intended and unintended changes to the speech signal. J Speech Lang Hear Res 62(7):2270–2279. https://doi.org/10.1044/2019_JSLHR-S-18-0408

Hickok G (2012) Computational neuroanatomy of speech production. Nat Rev Neurosci 13(2):135–145. https://doi.org/10.1038/nrn3158

Honda K (1983) Relationship between pitch control and vowel articulation. Haskins Laboratori Status Rep Speech Res 73:269–282

Houde JF, Chang EF (2015) The cortical computations underlying feedback control in vocal production. Curr Opin Neurobiol 33:174–181. https://doi.org/10.1016/j.conb.2015.04.006

Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82

Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: evidence from feedback perturbation. J Acoust Soc Am 108(3):1246–1251. https://doi.org/10.1121/1.1288414

Jones JA, Munhall KG (2002) The role of auditory feedback during phonation: studies of mandarin tone production. J Phon 30(3):303–320. https://doi.org/10.1006/jpho.2001.0160

Katseff S, Houde J, Johnson K (2012) Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? Lang Speech 55:295–308. https://doi.org/10.1177/0023830911417802

Kewley-Port D, Watson CS (1994) Formant-frequency discrimination for isolated English vowels. J Acoust Soc Am 95(1):485–496. https://doi.org/10.1121/1.410024

Lametti DR, Nasir SM, Ostry DJ (2012) Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. J Neurosci 32(27):9351–9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane H, Wozniak J, Matthies M, Svirsky M, Perkell J, O'Connell M, Manzella J (1997) Changes in sound pressure and fundamental frequency contours following changes in hearing status. J Acoust Soc Am 101(4):2244–2252. https://doi.org/10.1121/1.418245

Larson CR, Altman KW, Liu H, Hain TC (2008) Interactions between auditory and somatosensory feedback for voice F 0 control. Exp Brain Res 187(4):613–621

Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC (2001) Comparison of voice F0 responses to pitch-shift onset and offset conditions. J Acoust Soc Am 110(6):2845–2848. https://doi.org/10.1121/1.1417527

Larson CR, Burnett TA, Kiran S, Hain TC (2000) Effects of pitch-shift velocity on voice Fo responses. J Acoust Soc Am 107(1):559–564. https://doi.org/10.1121/1.428323

Lester-Smith RA, Daliri A, Enos N, Abur D, Lupiani AA, Letcher S, Stepp CE (2020) The relation of articulatory and vocal auditory-motor control in typical speakers. J Speech Lang Hear Res 63(11):3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am,= 49(2) Suppl 2:467+. https://www.ncbi.nlm.nih.gov/pubmed/5541744

Li W, Chen Z, Liu P, Zhang B, Huang D, Liu H (2013) Neurophysiological evidence of differential mechanisms involved in producing opposing and following responses to altered auditory feedback. Clin Neurophysiol 124(11):2161–2171. https://doi.org/10.1016/j.clinph.2013.04.340

Liu H, Auger J, Larson CR (2010a) Voice fundamental frequency modulates vocal response to pitch perturbations during english speech. J Acoust Soc Am 127(1):1–5. https://doi.org/10.1121/1.3263897

Liu H, Larson CR (2007) Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. J Acoust Soc Am 122(6):3671–3677. https://doi.org/10.1121/1.2800254

Liu H, Wang EQ, Chen Z, Liu P, Larson CR, Huang D (2010b) Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. J Acoust Soc Am 128(6):3739–3746. https://doi.org/10.1121/1.3500675

Liu H, Wang EQ, Chen Z, Liu P, Larson CR, Huang D (2010c) Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. J Acoust Soc Am 128(6):3739–3746

Liu P, Chen Z, Jones JA, Huang D, Liu H (2011) Auditory feedback control of vocal pitch during sustained vocalization: a cross-sectional study of adult aging. PLoS ONE 6(7):e22791. https://doi.org/10.1371/journal.pone.0022791

Loucks TM, Poletto CJ, Saxon KG, Ludlow CL (2005) Laryngeal muscle responses to mechanical displacement of the thyroid cartilage in humans. J Appl Physiol 99(3):922–930. https://doi.org/10.1152/japplphysiol.00402.2004

MacDonald EN, Munhall KG (2012) A preliminary study of individual responses to real-time pitch and formant perturbations. In: The Listening Talker: an interdisciplinary workshop on natural and synthetic modification of speech in response to listening conditions, pp 32–35

MacDonald EN, Purcell DW, Munhall KG (2011) Probing the independence of formant control using altered auditory feedback. J Acoust Soc Am 129(2):955–965. https://doi.org/10.1121/1.3531932

Macmillan NA, Creelman CD (2004) Detection theory: A user's guide. Psychology press

Martin CD, Niziolek CA, Dunabeitia JA, Perez A, Hernandez D, Carreiras M, Houde JF (2018) Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. Front Hum Neurosci 12:91. https://doi.org/10.3389/fnhum.2018.00091

McGill WJ, Goldberg JP (1968) Pure-tone intensity discrimination and energy detection. J Acoust Soc Am 44(2):576–581. https://doi.org/10.1121/1.1911123

Midi I, Dogan M, Koseoglu M, Can G, Sehitoglu M, Gunal D (2008) Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurol Scand 117(1):26–34

Miller HE, Guenther FH (2021) Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework. Aphasiology 35(4):424–441. https://doi.org/10.1080/02687038.2020.1765307

Miller N, Andrew S, Noble E, Walshe M (2011) Changing perceptions of self as a communicator in Parkinson's disease: a longitudinal follow-up study. Disabil Rehabil 33(3):204–210

Mitsuya T, MacDonald EN, Munhall KG, Purcell DW (2015) Formant compensation for auditory feedback with English vowels. J Acoust Soc Am 138(1):413–424. https://doi.org/10.1121/1.4923154

Mollaei F, Shiller DM, Baum SR, Gracco VL (2016) Sep 1). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. Brain Res 1646:269–277. https://doi.org/10.1016/j.brainres.2016.06.013

Mollaei F, Shiller DM, Baum SR, Gracco VL (2019) The relationship between speech perceptual discrimination and speech production in Parkinson's disease. J Speech Lang Hear Res 62(12):4256–4268. https://doi.org/10.1044/2019_JSLHR-S-18-0425

Munhall KG, MacDonald EN, Byrne SK, Johnsrude I (2009) Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. J Acoust Soc Am 125(1):384–390

Murray ESH, Stepp CE (2020) Relationships between vocal pitch perception and production: a developmental perspective. Sci Rep 10(1):1–10

Nasir SM, Ostry DJ (2006) Somatosensory precision in speech production. Curr Biol 16(19):1918–1923. https://doi.org/10.1016/j.cub.2006.07.069

Parrell B, Agnew Z, Nagarajan S, Houde J, Ivry RB (2017) Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. J Neurosci 37(38):9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell B, Ramanarayanan V, Nagarajan S, Houde J (2019) The FACTS model of speech motor control: fusing state estimation and task-based control. PLoS Comput Biol 15(9):e1007321. https://doi.org/10.1371/journal.pcbi.1007321

Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, Paul D, Svec JG, Hillman R (2018) Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of Vocal Function. Am J Speech Lang Pathol 27(3):887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Patel S, Nishimura C, Lodhavia A, Korzyukov O, Parkinson A, Robin DA, Larson CR (2014) Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback. J Acoust Soc Am 135(5):3036–3044. https://doi.org/10.1121/1.4870490

Perkell J, Matthies M, Lane H, Guenther F, Wilhelms-Tricarico R, Wozniak J, Guiod P (1997) Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. Speech Commun 22(2–3):227–250

Perkell, J. S. (1969) Physiology of speech production: Results and implications of a quantitative cineradiographic study MIT research monograph, No 53.

Perkell JS (2013) Five decades of research in speech motor control: what have we learned, and where should we go from here? J Speech Lang Hear Res 56(6):1857

Perkell JS, Cohen MH, Svirsky MA, Matthies ML, Garabieta I, Jackson MT (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. J Acoust Soc Am 92(6):3078–3096. https://doi.org/10.1121/1.404204

Perkell JS, Guenther FH, Lane H, Matthies ML, Perrier P, Vick J, Wilhelms-Tricarico R, Zandipour M (2000) A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. J Phon 28(3):233–272. https://doi.org/10.1006/jpho.2000.0116

Pisoni DB (1971) On the nature of categorical perception of speech sounds (Doctoral dissertation, ProQuest Information & Learning)

Purcell, D., & Munhall, K. (2008) Weighting of auditory feedback across the English vowel space. Proceedings of the... International Seminar on Speech Production

Purcell DW, Munhall KG (2006a) Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. J Acoust Soc Am 120(2):966–977. https://doi.org/10.1121/1.2217714

Purcell DW, Munhall KG (2006b) Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 119(4):2288–2297. https://doi.org/10.1121/1.2173514

Reilly KJ, Dougherty KE (2013) The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory

feedback. J Acoust Soc Am 134(2):1314–1323. https://doi.org/10.1121/1.4812763
Romo R, Hernandez A, Salinas E, Brody CD, Zainos A, Lemus L, de Lafuente V, Luna R (2002) From sensation to action. Behav Brain Res 135(1–2):105–118. https://doi.org/10.1016/s0166-4328(02)00161-4
Ryan TA, Joiner BL, Ryan BF (2014) MinitabTM. Statistics Reference Online, Wiley Online library, Wiley StatsRef
Sapir S (1978) The motions of the hyoid bone and thyroid cartilage during changes in the voice fundamental frequency: A cine-fluorographic pilot study. Doctoral dissertation, University of Washington
Sapir S, Baker KK, Larson CR, Ramig LO (2000) Short-latency changes in voice F0 and neck surface EMG induced by mechanical perturbations of the larynx during sustained vowel phonation. J Speech Lang Hear Res 43(1):268–276. https://doi.org/10.1044/jslhr.4301.268
Sapir S, Campbell C, Larson C (1981) Effect of geniohyoid, cricothyroid and sternothyroid muscle stimulation on voice fundamental frequency of electrically elicited phonation in rhesus macaque. Laryngoscope 91(3):457–468. https://doi.org/10.1288/00005537-198103000-00017
Scheerer NE, Jones JA (2012) The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. Neurosci Lett 529(2):128–132. https://doi.org/10.1016/j.neulet.2012.09.012
Skodda S, Grönheit W, Schlegel U (2012) Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. PLoS ONE 7(2):e32132
Smith DJ, Stepp C, Guenther FH, Kearney E (2020) Contributions of auditory and somatosensory feedback to vocal motor control. J Speech Lang Hear Res 63(7):2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296
Svirsky MA, Lane H, Perkell JS, Wozniak J (1992) Effects of short-term auditory deprivation on speech production in adult cochlear implant users. J Acoust Soc Am 92(3):1284–1300. https://doi.org/10.1121/1.403923
Terband H, Maassen B, Guenther FH, Brumberg J (2014) Auditory–motor interactions in pediatric motor speech disorders: neurocomputational modeling of disordered development. J Commun Disord 47:17–33
Tourville JA, Guenther FH (2011) The DIVA model: A neural theory of speech acquisition and production. Lang Cogn Process 26(7):952–981. https://doi.org/10.1080/01690960903498424
Tourville JA, Reilly KJ, Guenther FH (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 VM, Perkell JS, Guenther FH (2007) Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. J Acoust Soc Am 122(4):2306–2319. https://doi.org/10.1121/1.2773966
Weerathunge HR, Abur D, Enos NM, Brown KM, Stepp CE (2020) Auditory-motor perturbations of voice fundamental frequency: feedback delay and amplification. J Speech Lang Hear Res 63(9):2846–2860. https://doi.org/10.1044/2020_JSLHR-19-00407
Weerathunge HR, Alzamendi GA, Cler GJ, Guenther FH, Cara E, Stepp CE, Zañartu M (in press). LaDIVA: A neurocomputational model providing laryngeal motor control for speech acquisition and production. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1010159
Weerathunge HR, Tomassi NE, Stepp CE (2022) What can altered auditory feedback paradigms tell us about vocal motor control in individuals with voice disorders? Perspectives of the ASHA Special Interest Groups 7(3):959–976. https://doi.org/10.1044/2022_PERSP-21-00195