ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Assessing Ecologically Valid Methods of Auditory Feedback Measurement in Individuals With Typical Speech

Nicole E. Tomassi,[a,b] Hasini R. Weerathunge,[b,c] Megan R. Cushman,[b] Jason W. Bohland,[d] and Cara E. Stepp[a,b,c,e]

[a] Graduate Program for Neuroscience, Boston University, MA [b] Department of Speech, Language & Hearing Sciences, Boston University, MA [c] Department of Biomedical Engineering, Boston University, MA [d] Department of Communication Science and Disorders, University of Pittsburgh, PA [e] Department of Otolaryngology—Head & Neck Surgery, Boston University School of Medicine, MA

ARTICLE INFO

Article History:
Received July 7, 2021
Revision received September 15, 2021
Accepted September 16, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Kate Bunton

https://doi.org/10.1044/2021_JSLHR-21-00377

ABSTRACT

Purpose: Auditory feedback is thought to contribute to the online control of speech production. Yet, the standard method of estimating auditory feedback control (i.e., reflexive responses to auditory–motor perturbations), although sound, requires specialized instrumentation, meticulous calibration, unnatural tasks, and specific acoustic environments. The purpose of this study was to explore more ecologically valid features of speech production to determine their relationships with auditory feedback mechanisms.
Method: Two previously proposed measures of within-utterance variability (centering and baseline variability) were compared with reflexive response magnitudes in 30 adults with typical speech. These three measures were estimated for both the laryngeal and articulatory subsystems of speech.
Results: Regardless of the speech subsystem, neither centering nor baseline variability was shown to be related to reflexive response magnitudes. Likewise, no relationships were found between centering and baseline variability.
Conclusions: Despite previous suggestions that centering and baseline variability may be related to auditory feedback mechanisms, this study did not support these assertions. However, the detection of such relationships may have required a larger degree of variability in responses, relative to that found in those with typical speech. Future research on these relationships is warranted in populations with more heterogeneous responses, such as children or clinical populations.
Supplemental Material: https://doi.org/10.23641/asha.17330546

In everyday communication, speech is produced through a process of learning and monitoring to ensure that the intended message is being conveyed to the listener. Auditory feedback is an important but difficult-to-study contributor to speech. Thus, the purpose of this study was to examine the ecological validity of various, previously proposed measures of auditory feedback. An abundance of literature suggests a critical role for auditory feedback in helping to learn precise speech motor control and in maintaining the neural representations used to produce speech (e.g., Levitt et al., 1980; Oller & Eilers, 1988; C. R. Smith, 1975). For instance, speech development is impaired in infants with severe-to-profound hearing loss (e.g., Koopmans-van Beinum et al., 2001; Oller & Eilers, 1988). However, when hearing loss occurs postlingually, speech remains largely intelligible (e.g., Cowie & Douglas-Cowie, 1992; Lane et al., 1997). This difference in the impact of the presence of auditory feedback as a function of development is further informed by studies of partially restored hearing using cochlear implants (e.g., Banfai et al., 1984; Lane et al., 1991; Leder et al., 1986; Perkell et al., 1992). When auditory feedback was reintroduced in adults with late-onset hearing loss, gains in speech production were

Correspondence to Nicole E. Tomassi: ntomassi@bu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

observed, whereas adults with prelingual deafness did not experience the same improvements (e.g., Perkell et al., 1992). These seemingly disparate findings have led to efforts to model speech motor control in order to define the role(s) of auditory feedback more precisely in speech production.

Contemporary models of speech motor control have proposed various mechanisms to generate motor commands and ensure the realization of speech goals (Houde & Nagarajan, 2011; Parrell et al., 2019; Tourville & Guenther, 2011). Although models differ in their execution, there is an overall consensus that separate but parallel processes exist for speech production. Previously learned programs are used to send motor commands to speech articulators, whereas feedback systems are responsible for detecting and correcting errors between expected and perceived sensory signals (Houde & Nagarajan, 2011; Parrell et al., 2019; Tourville & Guenther, 2011). According to the Directions Into Velocities of Articulators (DIVA) model, both feedforward and feedback control subsystems are constructed and refined during development as relationships between motor commands and sensory consequences are defined and auditory targets are learned (Tourville & Guenther, 2011). The feedforward system, when mature, is responsible for sending motor commands fluently and quickly for well-learned sound sequences without the use of feedback. During speech acquisition or for novel sound sequences, feedback systems play a primary role by correcting for errors and refining learned motor commands within the feedforward system until they are able to drive intelligible, error-free speech production (Ghosh et al., 2008; Tourville & Guenther, 2011). According to the DIVA model, once the feedforward system has matured for common sound sequences, the feedback system is only utilized when an unexpected sensory response occurs, explaining why adults with postlingual deafness do not experience problems with speech to the same extent as those with prelingual deafness. In this theoretical framework, inducing an auditory error by manipulating the feedback a speaker receives should engage the auditory feedback control system and elicit error correction mechanisms. Thus, real-time experimental perturbations of the auditory system have become a useful and common approach to study auditory feedback control in individuals with typical hearing.

In the past few decades, technological development has allowed researchers to study the real-time effects of spectrally altering auditory feedback in controlled laboratory experiments, resulting in improved specificity in the understanding of auditory feedback mechanisms necessary to develop and evaluate models (Burnett et al., 1998; Elman, 1981; Houde & Jordan, 1998; Larson et al., 2001). Two types of perturbation protocols have emerged and have become the established methods for investigating the effects of auditory feedback control in the vocal and articulatory subsystems. Adaptation paradigms target learning within the feedforward system and involve predictable and sustained perturbations to auditory feedback, typically resulting in gradual adjustments of articulatory or vocal output as the feedforward controller "adapts" its commands in response to errors detected in the feedback control system, (partially) compensating for the perturbation (e.g., Houde & Jordan, 1998; Jones & Munhall, 2000; Patel et al., 2011). In contrast, reflexive paradigms target the online actions of the auditory feedback control system and use unpredictable and sudden perturbations to auditory feedback, resulting in rapid, real-time adjustments of articulatory or vocal output (e.g., Burnett et al., 1998; Tourville et al., 2008). Most participants tend to produce partial "compensatory" responses, which oppose the direction of the perturbation (e.g., Burnett et al., 1998; Lester-Smith et al., 2020; Tourville et al., 2008), thus steering auditory feedback closer to the learned auditory target(s). Reflexive perturbation experiments are currently the widely approved methodology for assessing auditory feedback control and have been an essential part of informing models of speech production.

Although current methods of measuring auditory feedback control through reflexive paradigms can provide valuable insight, they are often not ecologically valid, and technical limitations exist. These experiments typically require specialized equipment and software, meticulous calibration, unnatural tasks, and specific acoustic environments, which prevent these techniques from being widely used in clinical settings and limit the extent to which we can generalize results to other contexts. For these reasons, researchers have been exploring more ecologically valid methods to study auditory feedback control in speech production. In the laryngeal subsystem, researchers have proposed within-trial variability of voice fundamental frequency (f_o) during speech tasks as a potential window into auditory feedback control (Scheerer & Jones, 2012; Scheerer et al., 2013, 2016). In the articulatory subsystem, a new method of "vowel centering," which measures within-utterance change in formant frequencies, has been proposed as a method to assess auditory feedback control (Niziolek & Guenther, 2013; Niziolek & Kiran, 2018; Niziolek et al., 2013, 2015).

Variance of voice f_o in the period prior to a perturbation, termed the *baseline*, has been studied in reflexive pitch perturbation tasks, with inconclusive results to date. The theory behind this measure suggests that individuals with more variability during this initial time period will produce less consistent vocal output (due to weak or imprecise commands from the feedforward subsystem) and, thus, will rely more heavily on auditory feedback. This should, in turn, result in increased sensitivity to auditory–motor perturbations and larger compensatory responses. Some studies have demonstrated a strong positive correlation between vocal response magnitudes to unexpected pitch shifts and vocal baseline variability in children, aged 4–17 years, and adults, aged 18–36 years (Scheerer &

Jones, 2012; Scheerer et al., 2013). Other studies have found similar results but only in children aged 6–11 years and not in adults aged 18–28 years (Heller Murray & Stepp, 2020; Rathna Kumar et al., 2013). However, Scheerer et al. (2016) did not find this relationship in either adults or children. The relationship between vocal response magnitudes to unexpected pitch shifts and vocal baseline variability has also been explored in persons with Parkinson's disease. Both Huang et al. (2016) and Chen et al. (2013) found positive correlations between vocal baseline variability and vocal response magnitude in persons with Parkinson's disease, but not in their associated control groups. In the articulatory subsystem, the correlation between reflexive responses to formant frequency perturbations and formant baseline variability has not been reported but could be useful in exploring new, more ecologically valid measures of auditory feedback control of articulation.

In the articulatory subsystem, some recent studies have used acoustic variance, measured by within-utterance changes in vowel formants, to study auditory error correction in a naturalistic setting (Niziolek et al., 2013, 2015). In these studies, the change in vowel formants between the beginning and the middle of each utterance was measured and compared with the median of all utterances to assess "centering behavior." Vowel formants tended to move inward, in comparison to the median of all utterances, from the periphery of the formant distribution toward the center of the formant distribution over the duration of the utterance (Bakst & Niziolek, 2019; Niziolek & Kiran, 2018; Niziolek et al., 2013, 2015). This phenomenon has been interpreted as a compensatory response to auditory feedback, based on work by Niziolek et al. (2015), which measured centering with and without the presence of auditory masking noise. Although centering was observed both with and without masking, the magnitude of centering was reduced when auditory masking was used, compared with the condition without masking (Niziolek et al., 2015). This work provided preliminary evidence that centering may be the by-product of auditory feedback control (i.e., vowels that initiate further from the distributional center evoke the feedback controller for error correction). However, the mature speech motor control system has been thought to rely on feedforward mechanisms, unless auditory feedback is perturbed (Tourville & Guenther, 2011). In a recent study, Niziolek and Parrell (2021) compared centering to formant perturbation responses in the same individuals and did not find a relationship. However, the study was unlike typical perturbation experiments in that the purpose was to examine if altering the perceived reliability of auditory feedback affects participants' compensation to auditory perturbations. The study altered the reliability of auditory feedback through slight manipulations of feedback prior to giving typical-sized perturbations of formant frequencies. Thus, because the auditory reliability was manipulated, adding a confounding variable, it does not provide sufficient evidence to detail a lack of relationship between centering and typical formant perturbation responses. Centering has been applied in translational work in second-language learners and individuals with aphasia (Bakst & Niziolek, 2019; Niziolek & Kiran, 2018) and could be a possible ecological alternative to perturbation methods for studying auditory feedback control. However, centering has not yet been adequately compared with typical reflexive auditory feedback responses, nor has it been examined in the laryngeal subsystem. Thus, to better define how centering behavior might index auditory feedback control, it is important to determine its relationship with typical reflexive responses. Additionally, measuring centering behavior in the same speakers in both the laryngeal and articulatory subsystems can help determine if the observed effects generalize.

Although baseline variability and vowel centering have been explored independently in the laryngeal and articulatory subsystems, they both hinge on measurements of production changes in speech. Thus, although they are formulated differently, both measures are thought to be related to speakers' use of auditory feedback and, thus, could be related to each other. If baseline variability and centering are indeed related in individual participants, this would promote overall understanding and potentially motivate the standardization of methods that are most useful for understanding auditory feedback control for use in future translational research.

The purpose of this study was to determine if two ecologically motivated measures of speech production are related to auditory feedback control mechanisms for voice and articulation. We measured baseline variability and centering in the laryngeal subsystem using voice f_o and in the articulatory subsystem using vowel formant contours. Baseline variability and centering in both subsystems were compared with reflexive responses to sudden perturbations of f_o and the first vowel formant (F_1). We hypothesized that auditory feedback can be studied behaviorally by measuring ways that speakers naturally alter their speech. On the basis of this hypothesis, we predicted that the two measures of within-utterance changes (baseline variability and centering) would be related to the reflexive response to corresponding acoustic perturbations in each subsystem and that these two more ecologically valid measures would be related to each other.

Method

Participants

Participants were nonsmoking native English speakers aged 18–24 years (15 cisgender males, 14 cisgender females, and one nonbinary woman; M_{age} = 21.1 years, SD = 1.9). They self-reported no history of speech, language, hearing, or neurological disorders and no prior training in singing. The participants provided written consent as per the Boston University Institutional Review Board. All participants passed a hearing screening at levels of 25 dB HL at the frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz (American Speech-Language-Hearing Association, 2020). All participants completed a Voice-Related Quality of Life questionnaire with scores ranging from 10 to 22 (M = 12.1, SD = 2.5), indicating that they perceived their voice as having a fairly low impact on their day-to-day activities.

Instrumentation

All tasks for this study were recorded in a sound-treated booth. A Shure MX153 omnidirectional microphone was positioned 7 cm from the mouth at a 45° angle from the midline of the face (Patel et al., 2018). The microphone signal was amplified using an RME QuadMic II microphone preamplifier. During f_o perturbations, the microphone signal was digitized using an RME Fireface UCX sound card with 32-bit resolution and a sampling rate of 44100 Hz. During F_1 perturbations, the microphone signal was digitized using a MOTU UltraLite-mk3 Hybrid sound card with 32-bit resolution and a sampling rate of 44100 Hz. To accomplish f_o perturbations, an Eventide Eclipse V4 Harmonizer was used, which induces a delay of 11.1 ms (SD = 7.5) in the signal path (Heller Murray et al., 2019). To accomplish F_1 perturbations, Audapter (Cai et al., 2008), a software package for a configurable, real-time manipulation of acoustic parameters of speech, was used in MATLAB (Release 2018; MathWorks), which has a 20-ms delay[1] in our system. Sennheiser HD 280 Pro over-ear headphones were used to provide participants with amplified auditory feedback of their own speech productions. To overcome bone-conducted auditory feedback while minimizing unnecessarily high sound pressure levels, the auditory feedback was amplified by 5 dB relative to the sound pressure level at the microphone (Weerathunge et al., 2020). This gain level was used for both centering and reflexive tasks.

Experimental Design

This study consisted of multiple speech tasks using unpredictable and sudden perturbations of f_o and F_1 auditory feedback, referred to as reflexive tasks, as well as tasks with no perturbations, referred to as centering tasks. Centering tasks were always presented first to ensure that no lingering effects from perturbations in the reflexive tasks affected them, followed by F_1 reflexive tasks and, lastly, f_o reflexive tasks. Reflexive tasks were used to measure auditory feedback control via reflexive responses and baseline variability. These tasks comprised four blocks, two of which involved sustained reflexive perturbations of f_o auditory feedback, and the other two involved sustained reflexive perturbations of F_1 auditory feedback. In each block, the participants were exposed to the perturbed feedback in 25% of trials[2] and the unperturbed feedback in the other 75% of trials (perturbed trials were pseudorandomized across the experiment). The intertrial interval was 1, 1.5, or 2 s, determined randomly. In between each block, participants rested for 2 min. Centering tasks were used to measure within-utterance changes in f_o and formants and consisted of one block of trials in which participants received unperturbed feedback of their voice. In both the reflexive and centering tasks, participants were asked to produce approximately 1-s productions of the word “id” in response to associated visual cues on a computer screen. These somewhat prolonged productions were used to ensure that both feedforward and later feedback responses were captured in the resultant acoustic signal. This word was chosen to (a) elicit the vowel /I/, as a previous study in our group showed participants reliably responded to upward perturbations of this vowel (Lester-Smith et al., 2020), and (b) align the onset of phonation to the onset of the vowel.

f_o Reflexive Task

Prior to starting the f_o reflexive task, participants practiced 12 trials, three of which were randomly shifted to familiarize themselves with f_o perturbations. Participants then completed two blocks of 72 trials, with 18 trials in which f_o auditory feedback was perturbed per block, resulting in 36 total perturbed trials. In these 36 trials, participants received auditory feedback shifted downward by 100 cents. The perturbation onset was jittered by 200–500 ms after the onset of phonation, and the perturbation was sustained throughout the rest of the production.

F_1 Reflexive Task

The formant reflexive task followed a similar protocol as the f_o reflexive task. However, in the formant reflexive paradigm, the perturbation was applied to F_1 instead of voice f_o. Similar to the f_o reflexive task, a total of 36 trials were perturbed. F_1 in these trials was shifted upward by 30%, distorting the perception of the vowel /I/ toward the vowel /ɛ/ (i.e., “Ed”).

[1]Kim et al. (2020) reported delays as short as 10 ms and as long as 45 ms, with most systems ranging between 10 and 15 ms. We measured a delay of 20 ms during formant perturbations in our system.

[2]Perturbed feedback was given in only 25% of trials to avoid learning across trials and to avoid adaptation.

Centering Task

In order to determine centering behavior in both subsystems, f_o contours and formant trajectories were extracted from acoustic recordings (see below). Participants practiced 12 unperturbed trials and then completed a total of 108 unperturbed trials of the word "id" in a single block.

Data Analysis

For all acoustic analyses, voice f_o was estimated every 5 ms using Praat (Versions 5 and 6.0.40; Boersma & Weenink, 2016). F_1 and F_2 were estimated every 2 ms using Audapter. All trials were manually inspected and corrected for mistracking issues.

Reflexive Responses

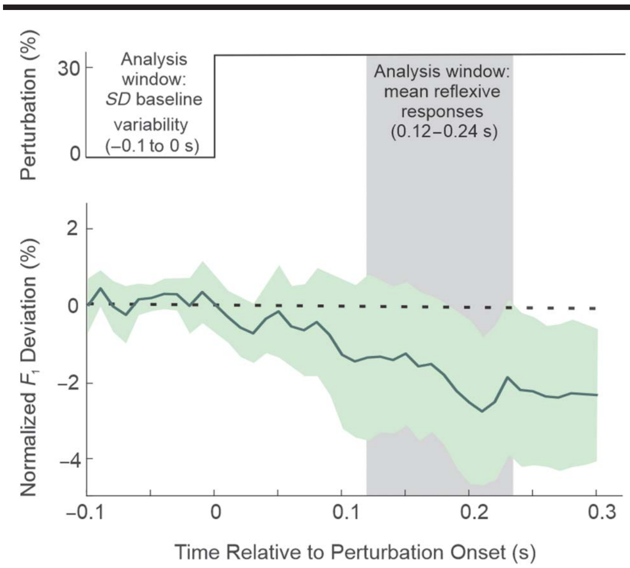
Reflexive response analysis was performed using custom Praat and MATLAB scripts. An illustrative example from one task (F_1 perturbation) is shown in Figure 1. For each trial, phonation onset, phonation offset, and perturbation onset were manually determined using a custom MATLAB graphical user interface. A wideband spectrogram of the auditory feedback to the headphones was used to detect perturbation onset. Trials were time-aligned to the start of the perturbation, and the baseline was defined as the 100 ms prior to perturbation onset. Individual trials' f_o contours were converted to cents relative to the mean of the baseline region (f_o baseline) for that trial using the following equation:

$$\text{Normalized } f_o \text{ deviation (cents)} = 1200 * \log_2 \frac{f_o}{f_o \text{ baseline}}. \quad (1)$$

F_1 was normalized, in percentages, relative to the mean of the baseline (F_1 baseline) using the following equation:

$$\text{Normalized formant deviation (\%)} = 100 * \left(\frac{F_1 - F_1 \text{ baseline}}{F_1 \text{ baseline}} \right). \quad (2)$$

Figure 1. The lower panel shows an example of one participant's averaged data from the first formant (F_1) reflexive task. The shading indicates the 95% confidence interval. The mean during the reflexive response window (highlighted in gray) was taken during 0.12–0.24 s for both F_1 and fundamental frequency reflexive tasks. The standard deviation (*SD*) was taken during the 0.1 s prior to perturbation onset and represented the baseline variability. The perturbation is schematized in the upper panel.



To avoid washout effects from perturbed trials, the two control trials following a perturbed trial were removed from analysis in each block. The remaining 36 control trials were averaged for each participant. The average of the control trials was then subtracted from the average of the 36 F_1- and f_o-perturbed trials.

Responses to reflexive paradigms to f_o typically occur with a latency of 100–120 ms (Burnett et al., 1998), and voluntary responses are thought to begin after 200–400 ms (Hain et al., 2000; H. Liu & Larson, 2007). Thus, reflexive responses in both subsystems were defined as the mean of the response during 120–240 ms from perturbation onset. The mean, rather than the peak, during this time window was used to quantify the reflexive response due to the sustained shift used in this paradigm. In other paradigms in which there is an onset and offset to the shift, there is a clearer and more defined peak available for quantification of the response (Burnett et al., 1998; Larson et al., 2001). However, because sustained perturbations often result in a plateaued response, the mean of this time interval is more effective at capturing the reflexive response. It is common for reflexive responses to be evaluated in one of two ways, either by (a) measuring the average response magnitude regardless of response direction, which is more common in formant perturbation studies, or by (b) sorting and averaging the responses by direction, opposing the shift (compensatory) or following the shift, which is more common in f_o perturbation studies (Behroozmand et al., 2012). Due to the lack of compelling evidence for only using compensatory trials and our interest in the variability within responses, we did not separate between following and compensatory responses and included all trials in the analysis. Since the perturbations for the f_o and F_1 reflexive paradigms were applied in opposite directions, to avoid confusion, responses to the f_o reflexive paradigm were negated so that compensatory responses in both subsystems would be negative.

Baseline Variability

The standard deviation of f_o during the baseline period of each f_o reflexive trial was calculated after the conversion (see Equation 1 above) from hertz to cents. These values were then averaged to produce the average f_o baseline variability for each participant. Likewise, the standard deviation of F_1 during the baseline period of each formant reflexive trial was calculated after determining the normalized formant deviation from the mean and averaged to determine the F_1 baseline variability for each participant.

Centering

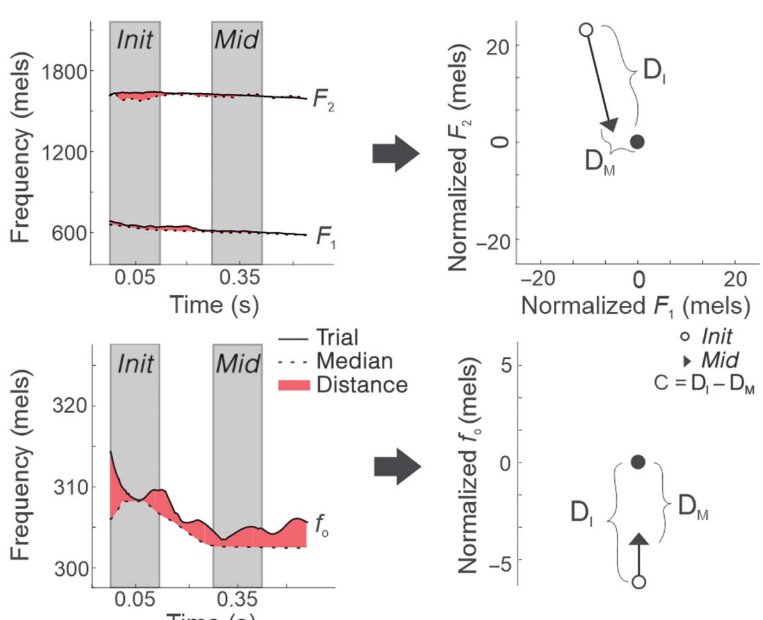
The calculation of centering behavior is schematized in Figure 2. For normalization purposes, f_o, F_1, and F_2 were converted to mels, a perceptually based logarithmic scale, as in Niziolek et al. (2015). The hertz-to-mels conversion was performed with the following equation:

$$\text{mels} = 2595 \ \log_{10}\left(1 + \frac{\text{Hz}}{700}\right). \quad (3)$$

To determine if there was movement toward the center of the f_o and formant distribution space over the course of the utterance, two time windows for analysis were established at the onset and "middle" of the utterance (Bakst & Niziolek, 2019; Niziolek & Kiran, 2018; Niziolek et al., 2013, 2015). The f_o and formant contours were averaged in the first 100 ms of each utterance (f_{oI} and F_{1I}, F_{2I}) and the 300- to 400-ms time frame of each utterance (f_{oM} and F_{1M}, F_{2M}). These time windows were adapted and modified from previous centering studies, which used the first 100 ms and middle 50% of the utterance (Bakst & Niziolek, 2019; Niziolek & Kiran, 2018; Niziolek et al., 2013, 2015), to account for the longer trial length in this study and to ensure that the time windows had equal duration.

In both time windows, the participant-wise medians $\left(\hat{f}_{oI}, \hat{f}_{oM}, \hat{F}_{1I}, \hat{F}_{2I}, \hat{F}_{1M}, \text{and } \hat{F}_{2M}\right)$ were calculated. Trials were then categorized based on their initial distances from the median, that is, $D_I(f_o)$ and D_I(Formants), using the initial time window. Trials were defined as "peripheral" or "center" trials based on the terciles with the greatest and smallest distance, respectively, of each participant's trial distribution. The distances from the median value in the middle time window, that is, $D_M(f_o)$ and D_M(Formants), were also calculated. For f_o centering, the

Figure 2. Schematic representation of centering analyses. The left panels represent the data (in mels) of one trial in a single participant. The black lines represent the participant's acoustic data—either first and second vowel formants (F_1, F_2; upper) or voice fundamental frequency (f_o; lower). The dotted lines represent the participant's median acoustic data across all trials. The red shading indicates the difference between the participant's single-trial production and the participant's median production across trials. The vertical, gray-shaded regions indicate the time windows in which the analyses took place. The initial (init) time window was the first 0.1 s following voice onset, and the middle (mid) time window was 0.3–0.4 s. The right panels represent the analysis of centering in F_1 and F_2 (upper) and f_o (lower). The open circles represent the average acoustic data of the one trial during the initial time window, and the arrowheads represent the average acoustic data of the one trial during the middle time window. The filled circles represent the participant-wise median from both time windows, meaning the open circles and arrowheads are normalized to their respective medians. D_I and D_M represent the distance from the initial and middle time windows, respectively, to the median. Movement toward the filled circles illustrates centering behavior (C).



initial and middle utterance distances from the median were calculated in one-dimensional space as follows:

$$D_I(f_o) = f_{oI} - \hat{f}_{oI}, \quad (4)$$

$$D_M(f_o) = f_{oM} - \hat{f}_{oM}. \quad (5)$$

For formant centering, the initial and middle utterance distances from the median were calculated as Euclidean distances in two-dimensional formant space (F_1 and F_2) using the following formulas:

$$D_I(\text{Formants}) = \sqrt{\left(F_{1I} - \hat{F}_{1I}\right)^2 + \left(F_{2I} - \hat{F}_{2I}\right)^2}, \quad (6)$$

$$D_M(\text{Formants}) = \sqrt{\left(F_{1M} - \hat{F}_{1M}\right)^2 + \left(F_{2M} - \hat{F}_{2M}\right)^2}. \quad (7)$$

Centering (C) was calculated for all trials and defined for both f_o and formants as follows:

$$C = D_I - D_M. \quad (8)$$

Larger or positive values of C corresponded to stronger centering, whereas smaller or negative values of C corresponded to weaker centering. Since this behavior is thought to be corrective, it would be expected that the more deviated or "peripheral" trials would exhibit a greater magnitude of centering than the "center" trials that are closer to the target upon phonation onset (Niziolek et al., 2013, 2015). Thus, centering responses from the 36 "peripheral" trials only were used in the subsequent statistical analysis.

Statistical Analysis

All statistical analyses were performed in Minitab statistical software. Correlation analysis was performed, across participants, on both vocal and articulatory data to characterize relationships between the proposed ecologically valid methods and current gold-standard methods of measuring auditory feedback. Within subsystems, reflexive responses, baseline variability, and centering were compared with each other using Pearson correlation coefficients. To ensure that the results were not impacted by the logarithmic nature of most units, correlation analysis was also performed on transformed data. This entailed converting all nonlinear data (i.e., f_o reflexive responses and baseline variability in cents and all centering data in mels) back to the original units (Hz) for comparison to the already linear data (F_1 reflexive responses and baseline variability). An α of .05 was used as a threshold for significance testing. As an additional exploratory analysis, we performed a Pearson correlation analysis on all measures (reflexive responses, baseline variabilities, and centering magnitudes) across subsystems (voice and articulation) to determine if responses were generalizable within individual participants.

Results

In both the laryngeal and articulatory subsystems, participants tended to produce compensatory reflexive responses. Because all f_o responses to "shift-down" perturbations were negated, compensatory values are indicated by negative values. The f_o reflexive responses were −7.5 cents on average (*SD* = 11.8), changing their f_o in the opposite direction of the perturbation (see Figure 3, upper panel); however, average responses across participants ranged from −35.5 to 12.9 cents, with 76.7% of these responses being compensatory. Similarly, 70.0% of F_1 reflexive responses across participants were compensatory, with an average reflexive response of −0.61% (*SD* = 0.99%) and a range of −2.78% to 0.36% (see Figure 3, lower panel). Individual participant behavior for the reflexive paradigms can be seen in Supplemental Material S1 for the laryngeal subsystem and in Supplemental Material S2 for the articulatory subsystem.

The mean f_o baseline variability was 2.97 cents (*SD* = 2.12). Across participants, the values ranged from 0.71 to 9.66 cents. The mean F_1 baseline variability was 0.38% (*SD* = 0.36%), with participant values ranging from 0.06% to 1.83%.

Centering behavior was observed in the articulatory subsystem for all participants and in all but one participant in the laryngeal subsystem. Thus, for most participants, average centering values were greater than zero, indicating that participants moved their f_o and formants toward their own median values over the duration of each utterance (see Figure 4). Participants exhibited an average f_o centering of 1.87 mels (*SD* = 1.74), with a range from −0.13 to 8.42 mels. Participants exhibited an average formant centering of 26.32 mels (*SD* = 19.60), with a range from 3.51 to 90.45 mels. Individual participant behavior for the centering tasks can be seen in Supplemental Material S3 for the laryngeal subsystem and in Supplemental Material S4 for the articulatory subsystem.

In the laryngeal subsystem, neither f_o baseline variability nor f_o centering was statistically significantly correlated with reflexive responses (see Table 1). Furthermore, f_o baseline variability and f_o centering were also not statistically significantly correlated with each other. Similarly, in the articulatory subsystem, neither formant baseline variability nor formant centering was statistically significantly correlated with F_1 reflexive responses (see Table 1). Moreover, formant baseline variability and formant centering were not statistically significantly correlated with each other (see Table 1). The results of these correlation analyses on the transformed data also yielded no significant correlations. The results of our exploratory analysis

Figure 3. Average reflexive responses for all participants. The average of responses to a 100-cent fundamental frequency (f_o) shift-down is shown (upper), with shading representing the 95% confidence interval. The average of responses to a 30% shift-up in the first formant (F_1) is shown (lower), with shading representing the 95% confidence interval. Here, f_o reflexive responses were negated for ease of viewing compensatory responses in both subsystems. Baseline variability (not shown) was calculated as the standard deviation of the 0.1 s prior to perturbation onset. The mean f_o baseline variability was 2.97 cents (*SD* = 2.12). The mean F_1 baseline variability was 0.38% (*SD* = 0.36%).

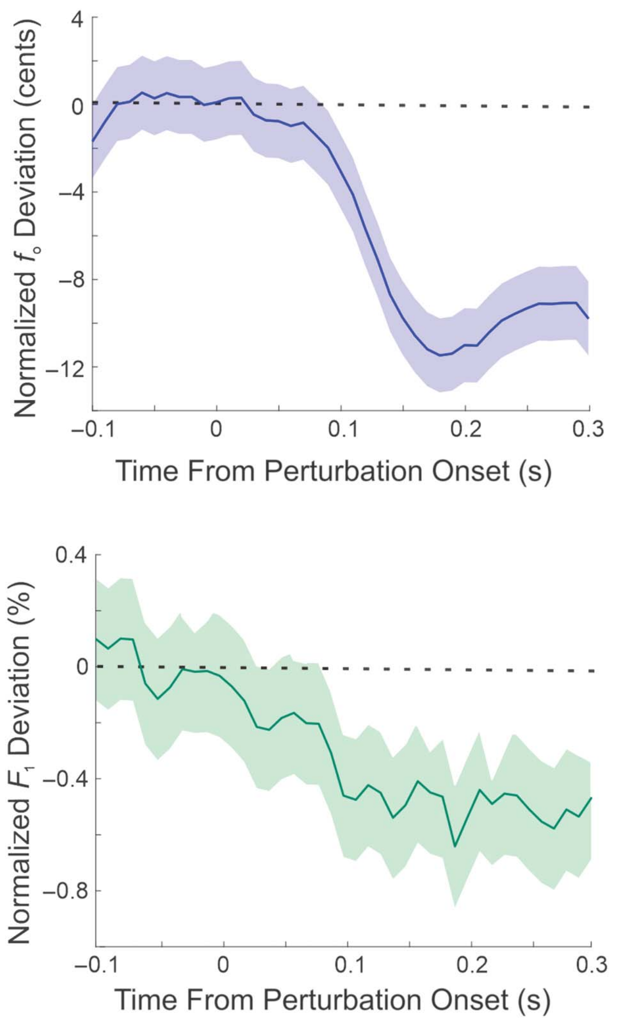
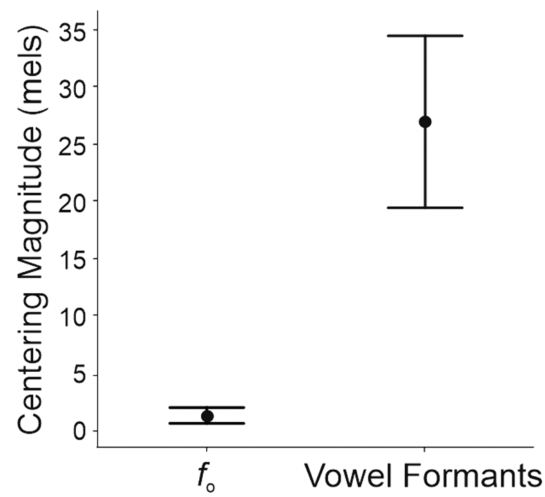


Figure 4. Centering responses (in mels) for both fundamental frequency (f_o) and vowel formants. Error bars represent 95% confidence intervals.



across subsystems were as follows: F_1 reflexive responses were not correlated with f_o reflexive responses ($r = -.24$, $p = .21$), F_1 baseline variability was not correlated with f_o baseline variability ($r = .15$, $p = .43$), and formant centering was not correlated with f_o centering ($r = .03$, $p = .86$).

Discussion

This study investigated two features of speech production and their relationship to auditory feedback mechanisms in the laryngeal and articulatory subsystems. We hypothesized that baseline variability and centering in both subsystems would be related to reflexive responses, which are considered the current gold standard for probing online auditory feedback control. However, the results were contrary to our hypotheses: We found no evidence that either baseline variability or centering in the laryngeal subsystem was statistically related to reflexive responses. In the articulatory subsystem, baseline variability and centering were also unrelated to reflexive responses. Thus, our results do not support the idea that baseline variability or centering responses measure auditory feedback control analogously to perturbation experiments in either subsystem. Additionally, we hypothesized that baseline variability and centering, both measures of temporal production variance, would be related; however, the measures were not statistically significantly correlated in either subsystem. Thus, our study does not support that these measures are capturing similar speech phenomena.

The reflexive responses documented in this study were consistent in direction and magnitude with those reported previously in the literature. Burnett et al. (1998) and P. Liu et al. (2011) reported magnitudes for reflexive responses (compensatory only) ranging between 2 and 100 cents. Using a more similar paradigm to this study, Lester-Smith et al. (2020) reported an average of all f_o reflexive responses to an upward 100-cent shift as −7.1 cents (*SD* = 6.0), with responses ranging from −17.9 to 9.9 cents. The average f_o reflexive response found in this study ($M = -7.5$ cents, $SD = 11.8$) falls within the range of these previous experiments. The f_o reflexive responses found in this study are most consistent with the values found in Lester-Smith et al., most likely due to the inclusion of both following and compensatory responses and the use of similar methodology. Average

Table 1. Pearson correlation coefficients for different features of speech (reflexive responses, baseline variability, and centering) in the laryngeal (voice fundamental frequency) and articulatory (vowel formants) subsystems.

	Voice fundamental frequency		
Features of speech	**Reflexive responses**	**Baseline variability**	**Centering**
Reflexive responses			
Baseline variability	−.05 (.81)		
Centering	.13 (.49)	.21 (.29)	
	Vowel formants		
Features of speech	**Reflexive responses**	**Baseline variability**	**Centering**
Reflexive responses			
Baseline variability	.06 (.73)		
Centering	−.06 (.74)	.05 (.76)	

Note. Here, *p* values are reported in parentheses. An α of .05 or less was used for significance testing.

F_1 reflexive responses from this study (M = −0.61%, SD = 0.99%) also aligned most similarly in direction and magnitude with values from Lester-Smith et al., which were reported as an average of −0.6% (SD = 1.3%) F_1 reflexive response to a 30% shift-up perturbation, with values ranging from −18% to 2.2%.

The baseline variability and centering values observed in this study were also compatible with relevant previous work. The average f_o baseline variability of this study (M = 2.97 cents, SD = 2.12) fell within the range of the values observed by previous researchers in adults with typical speech, who reported that values were between 0.1 and 8 cents (Chen et al., 2013; Heller Murray & Stepp, 2020; Huang et al., 2016; Scheerer & Jones, 2012). Because formant baseline variability has not been previously reported in this way in the literature, we did not have prior values with which to compare our findings (M = 0.38%, SD = 0.36%). Formant centering magnitudes from "peripheral" trials were similar to those in previous studies (Niziolek et al., 2013, 2015). Niziolek et al. (2013) reported an average centering response of 18 mels, and Niziolek et al. (2015) reported an average centering response of 45.2 mels. The average formant centering magnitude from this study was 26.3 mels, which falls within the range of the average centering values from the discussed studies. This study was the first to generate normative data for adults with typical speech of f_o centering, and thus, we had no values for comparison (M = 1.87 mels, SD = 1.74).

Relationship Between Baseline Variability and Reflexive Responses

Our results do not support a linear relationship between baseline variability and reflexive responses in the laryngeal or the articulatory subsystem, although previous research reported positive correlations between f_o reflexive responses and f_o baseline variability (Heller Murray & Stepp, 2020; Scheerer & Jones, 2012; Scheerer et al., 2013, 2016). The correlation between f_o reflexive responses and f_o baseline variability found in Heller Murray and Stepp (2020) was only observed in children, which could be, in part, due to the incomplete development of the speech motor control subsystems in this age range (6–11 years), causing more individual variability in which some relied more heavily on feedback. Other studies, including Chen et al. (2013) and Huang et al. (2016), found this same positive correlation in groups of speakers with Parkinson's disease but not in speakers with typical speech, likely due to the greater range of reflexive responses in the groups with Parkinson's disease. The range of within-group baseline variability in this study (0.7–9.6 cents) was only smaller than that for the group with Parkinson's disease in the work of Huang et al. and comparable to values found in the works of Chen et al., Heller Murray and Stepp, Scheerer et al. (2013, 2016), and Scheerer and Jones (2012). However, the range of reflexive response magnitudes in this study (from −35.5 to 12.9 cents) was smaller than that in the work of Scheerer and Jones as well as for the group with Parkinson's disease in the work of Huang et al. Thus, it is possible that the range of within-group variability plays a role in observing a relationship between the measures, thus more likely to be observed in populations with speech disorders or pediatric speakers with developing speech motor control systems.

Disparate experimental designs may also explain the discrepancy between findings in this study and findings in previous studies that reported a relationship between f_o reflexive responses and f_o baseline variability. For example, Scheerer and Jones (2012) and Scheerer et al. (2013) used a study design that differed in the number of perturbed trials. In Scheerer et al. (2013), participants were exposed to shifted feedback in 50% of trials, and in Scheerer and Jones, participants were exposed to shifted feedback in 100% of trials. These perturbation ratios differ drastically from this study in which 25% of the trials were perturbed and risk the possibility for sensorimotor learning by the participants. However, other studies that did not find a relationship between f_o reflexive responses and f_o baseline variability, such as Scheerer et al. (2016), Huang et al. (2016), and Chen et al. (2013), had perturbed trial counts comparable to those found in the works of Scheerer and Jones and Scheerer et al. (2013); thus, it is unclear how the number of perturbed trials may affect study outcomes.

This study was the first to explore a relationship between baseline variability and reflexive responses in the articulatory subsystem. However, formant baseline variability was not significantly correlated with F_1 reflexive responses. It could be the case, as discussed in relation to the laryngeal subsystem, that the variability of speakers with typical speech found in this study was not sufficient to detect a relationship if one exists, although it is notable that the observed correlation coefficient (r = .06) was close to zero. There are currently far fewer studies investigating F_1 reflexive responses than f_o reflexive responses to perturbations, so it is unclear whether interspeaker variability is related to these null findings.

Relationship Between Centering and Reflexive Responses

In this study, centering and reflexive responses were not linearly correlated in either subsystem. This suggests that centering may not be a reliable indicator of auditory feedback control. Previous research suggested that auditory feedback contributes to centering behavior in the articulatory subsystem but is not the sole driver (Niziolek et al., 2015), as formant centering was observed even with auditory masking. An alternative explanation for this observation could be the use of somatosensory feedback, as speech production relies on both auditory and somatosensory feedback (Houde & Jordan, 1998; Jones & Munhall, 2005; Tremblay et al., 2003).

Evidence for the role of somatosensory feedback in speech production, independent of auditory feedback, has been observed in both laryngeal and articulatory subsystems through the anesthetization of human vocal folds and the use of robotic devices to alter the jaw's motion path (Leonard & Ringel, 1979; Nasir & Ostry, 2006; Sorensen et al., 1980). Other studies have explored the interactions of auditory and somatosensory feedback systems for both laryngeal and articulatory control (Feng et al., 2011; Katseff et al., 2012; Lametti et al., 2012; Larson et al., 2008; D. J. Smith et al., 2020). One interpretation from the results of these studies suggests differing weights of auditory and somatosensory control and a possible individual preference for one control system over the other (Feng et al., 2011; Katseff et al., 2012; Lametti et al., 2012). This interpretation was based on observations that participants produce a different degree of compensation when either auditory or somatosensory feedback is perturbed and/or masked in contrast to when both control systems are available. This interpretation may explain why Niziolek et al. (2015) observed reduced centering responses when auditory feedback was masked. The definition of centering suggests that the change in within-utterance variability is a measure of compensation. Thus, when auditory feedback is masked, the contribution of auditory feedback is removed, leaving only somatosensory feedback, resulting in lowered compensation or centering; however, when auditory feedback is available, the centering response may be due to the combination of both subsystems. However, there is another methodological difference between this study and previous centering studies to consider, which is the use of a single word versus multiple words as stimuli. In prior centering studies (Niziolek et al., 2013, 2015), three words were used in prompts, as opposed to the single word used in this study. The use of multiple words could alter the participants' weighting of sensory feedback, encouraging greater acoustic distinction between productions. The use of a single word in this study may have resulted in a different weighting of sensory feedback and could be related to the study results.

This study isolated the potential role of auditory feedback in centering by using established techniques to measure auditory feedback control (reflexive perturbation paradigms) and directly comparing to centering responses in individual participants. No significant correlations were found between reflexive response magnitudes and centering values in either the laryngeal or the articulatory subsystem, suggesting that either auditory feedback control is not a significant contributor to centering or centering also measures additional effects that make its magnitude not sufficiently comparable to reflexive response magnitudes. This could again be due to the lack of sufficient variability in the study population to show a significant relationship if one were to exist, as discussed in relation to our findings for f_o baseline variability. However, this supposition can be informed by a recent study that similarly examined centering as an auditory feedback mechanism in a sample of individuals with cerebellar ataxia (Parrell et al., 2021). Individuals with cerebellar ataxia have been shown to display increased reliance on auditory feedback, producing larger compensatory responses in auditory perturbation experiments (Houde et al., 2019; Li et al., 2019; Parrell et al., 2017). Thus, if centering were a robust measure of auditory feedback control, it would seem likely that people with cerebellar ataxia would center more than typical speakers; however, that was not the case (Parrell et al., 2021). In addition, Parrell et al. (2021) utilized a masking condition, similar to Niziolek et al. (2015), but reported no difference in centering in conditions with and without auditory masking. Furthermore, Niziolek and Parrell (2021) similarly found no relationship between centering and responses to formant perturbations. Therefore, on the basis of the current findings and the extant literature, centering is unlikely to be a reliable measure of auditory feedback control.

Relationship Between Centering and Baseline Variability

Centering and baseline variability were not found to be correlated in this study. This study was the first in investigating the relationship between baseline variability and centering in both the laryngeal and articulatory subsystems. Niziolek and Parrell (2021) compared baseline variability in adults with typical speech (albeit with a different measurement approach) and centering in the articulatory subsystem and did not find a relationship between the measures. This study measured baseline variability in the articulatory subsystem analogously to how it has been measured in the laryngeal subsystem in previous studies (i.e., Heller Murray & Stepp, 2020; Scheerer & Jones, 2012; Scheerer et al., 2013, 2016). Specifically, baseline variability was defined as the standard deviation from the mean during the time period prior to perturbation. In contrast, Niziolek and Parrell defined baseline variability in the articulatory subsystem as the Euclidean distance from the median of each production in the initial time window of unperturbed trials. Conceptually, both definitions measure the distance from some middle value at the start of produced speech and, thus, are not likely to provide qualitatively different relationships between baseline variability and centering. Since no association between centering and baseline variability was observed using either definition, current evidence suggests these probes measure different aspects of speech production or are dominated by noise.

Because centering and baseline variability both measure variation in speech output, it would seem likely that they would be related (i.e., speakers with more variability would also exhibit more centering). Centering measures the change in within-utterance variability from the start to middle of a production and normalizes individual trial variability to the median across all trials. Core differences between centering and baseline variability are that (a) baseline variability only determines the initial variability of a production and (b) centering has a sign to indicate direction of movement toward or away from the median. For consistent speakers, the difference between the measures of baseline variability and centering could be small and, thus, may not be large enough to determine a relationship, as discussed above with the possible variability issues in the study sample of participants. However, for speakers in which the output is inconsistent from the start of the production, the difference between baseline variability and centering becomes critical in determining a relationship between the measures. A large centering response indicates that the speaker converges to more prototypical targets over the course of production, resulting in lower mid-utterance compared to early-utterance variability. If centering and baseline variability were related, larger values of baseline variability should result in larger decreases in variability over the utterance and increased centering magnitudes. Previous results investigating centering in individuals with cerebellar ataxia do not support this hypothesis and show that inconsistent speakers remain inconsistent across their productions (Parrell et al., 2021). Therefore, although baseline variability and centering seem similar conceptually, current data suggest that variability at the start of a production does not predict that productions will converge toward acoustic targets over the course of the utterance.

Exploratory Analysis: Relationships Between Subsystems

Our exploratory analysis found no relationships between any measures across subsystems. This is unsurprising due to the differences in the physiological nature and specific speech features captured by articulatory and laryngeal parameters. Voice f_o is commonly associated with prosody and typically changes slower than articulatory parameters (Lehiste, 1970; Perkell et al., 2007). Articulatory parameters are used to produce specific phonemic information and must change quickly during speech to maintain intelligibility and ensure the intended phonemes are conveyed. These time-varying differences between laryngeal and articulatory parameters have been interpreted in terms of feedback and feedforward control mechanisms (Kim & Max, 2014; Maas et al., 2015; Perkell, 2012; Perkell et al., 2007). It has been suggested that articulatory parameters rely more on feedforward control to maintain accurate phoneme productions than feedback control, consistent with the DIVA model and evidence from individuals with postlingual deafness who remain intelligible without auditory feedback (Cowie & Douglas-Cowie, 1992; Tourville & Guenther, 2011; Tourville et al., 2008). Moreover, articulatory parameters are less affected by changes to auditory feedback, relative to laryngeal parameters (Perkell, 2012; Perkell et al., 2007). This is most likely due to the fact that the movements for phonemes are too fast to be guided by online feedback detection, yet laryngeal parameters vary over longer time scales, suggesting that they may be more susceptible to feedback corrections. These differences in speech motor control between articulatory and laryngeal parameters are critical to consider when examining relationships of outcomes across subsystems.

Reflexive Task Paradigm

The reflexive task paradigm in this study was similar to those performed by our group and other research groups (Heller Murray & Stepp, 2020; Lester-Smith et al., 2020; Tourville et al., 2008) but differs from traditional "pitch-shift reflex" tasks (Burnett et al., 1998; Larson et al., 2001). The main difference between these paradigms is the use of a sustained perturbation after onset (as in this study) versus the application and removal of one or more perturbations over a specific time interval. Potential limitations with the design of the latter paradigm include the following: (a) If using multiple perturbations in one trial, aftereffects of one reflexive response can overlap with the baseline of an adjacent perturbation, and (b) the overall response to each traditional reflexive perturbation contains the response to a perturbation onset as well as the response to its offset. Perturbation duration is typically around 200 ms in the traditional reflexive paradigms, and the analysis includes measuring the peak magnitude of the resulting response. This poses a potential problem because the response may include a combination of the phasic and the prolonged response (Hain et al., 2000), which are involuntary and voluntary, respectively, as well as the perturbation offset response. The task design of this study minimizes these limitations by using one sustained perturbation in a trial, spacing out trials to ensure no adjacent trials are perturbed, and using the mean of an analysis window of 120–240 ms to isolate the phasic responses. Benefits of the current study design are that we avoid unintentional adaptation responses or aftereffects from previous trial responses and that we capture the extent of the involuntary response to most accurately isolate auditory feedback control.

Limitations and Future Directions

This study relies on the assumption that reflexive perturbation paradigms can be used to veridically measure auditory feedback control. Reflexive paradigms have become the gold standard for examining auditory feedback control because perturbing speech feedback directly induces auditory error and triggers a response in the feedback control system. Thus, the measured responses are directly linked to the change in auditory feedback. Although responses typically show that perturbations to auditory feedback result in (partial) compensation (opposing the perturbation), other responses, such as those that “follow” the perturbation, have also been observed (Behroozmand et al., 2012; Lester-Smith et al., 2020). To date, there has been no consensus explanation for these unexpected responses, and some researcher groups ignore them entirely (e.g., H. Liu et al., 2007; P. Liu et al., 2011; Scheerer et al., 2013). These atypical responses highlight individual variability in reflexive responses, which could be due, in part, to the large auditory errors induced by unnaturally perturbing feedback. These large errors may be treated differently than those naturally occurring during speech, with previous findings suggesting that the size of induced sensory errors affects how the system weighs those errors (Wei & Kording, 2009). Since the cause of the variability in responses is unclear, we included both compensatory and following responses in the estimated reflexive responses in this study.

In addition to the variability in responses, the variability in the methodology of reflexive paradigms is a factor to consider when discussing the validity and overall acceptance of these experiments. The methodological details of reflexive paradigms differ from group to group, including the proportion of trials that are perturbed, the magnitude and direction of perturbation, and the duration of perturbation. In this study, we adopted protocols previously used by our research group (Lester-Smith et al., 2020), but it is impossible to determine to what extent the null relationships observed between measures hinged on specific task parameters. Thus, although reflexive perturbation paradigms may still be considered the optimal approach to understanding auditory feedback control, further consideration of how to analyze different response types and optimize methodological details is warranted.

Reproducibility within individuals must also be discussed when attempting to find correlations between measures that are often variable. It may be the case that because reflexive responses vary by speaker across trials, they may be indicative of a state rather than a trait. If so, this would explain the lack of correlation between reflexive responses and other measures such as centering and baseline variability. Although there is some evidence that suggests responses to sensory feedback perturbations are specific to the individual, that is, weighting somatosensory feedback over auditory feedback (Lametti et al., 2012), it remains unclear whether these preferences are a trait and, thus, would be reproducible. Therefore, due to the lack of literature on the reproducibility of individual responses to feedback perturbations, the idea that reflexive responses may be a state rather than a trait of the individual is a limitation of this study.

When calculating centering in the laryngeal subsystem, a procedure that has not been performed in prior centering studies, it was assumed that the methods translated directly from the two-dimensional space of vowel formants to the one-directional space of voice f_o. It is possible, however, that the idea of within-utterance “targets” in the articulatory subsystem does not convey the same meaning in the laryngeal subsystem. Because pitch is a feature of speech that is typically controlled throughout running speech, it is reasonable to expect stable pitch “targets” that are controlled within each production. In the articulatory subsystem, acoustic targets (i.e., F_1 and F_2 values) change depending on the vowel to be produced, requiring frequent articulator movements to produce intelligible speech. In the laryngeal subsystem, f_o generally varies more slowly in natural speech to reflect intonational changes, which were minimal in the brief speech stimuli used in this study. It would be interesting to determine if there are time-varying f_o targets that speakers attempt to consistently produce during more natural and longer productions of running speech. In this case, future studies could explore centering-like behavior on a longer time scale in the laryngeal subsystem.

Although auditory feedback control—as indexed by reflexive perturbation responses—was not found to be related to centering, it would be interesting to study centering responses in relation to other feedback systems. The somatosensory feedback system plays an important role in the production of speech and might contribute to centering behavior. Future studies could use tactile perturbations in the laryngeal and/or articulatory subsystem, while controlling for auditory feedback, to compare directly to centering values.

Conclusions

In summary, this study found no evidence for linear relationships between baseline variability and reflexive responses to perturbations in the laryngeal or the articulatory subsystem. Additionally, centering was not related to reflexive responses in either subsystem. These findings indicate that the ecologically valid methods of interest in this study are not clearly related to auditory feedback control in either subsystem in adults with typical speech. These results may be limited by the low amount of within-group variability found in the responses of adults with mature speech motor control systems, which may limit generalization of these findings. Further investigation of the relationship between these different measures in other more variable populations, including children and individuals with speech disorders, is warranted.

Acknowledgments

This study was funded by National Institute on Deafness and Other Communication Disorders Grants DC016270 (Cara E. Stepp and Frank H. Guenther), DC015446 (Robert E. Hillman), and DC013017 (Christopher A. Moore and Cara E. Stepp). The authors would like to recognize Anton Dolling for his efforts in recruiting participants for this study.

References

American Speech-Language-Hearing Association. (2020). *Adult hearing screening* [Practice portal]. https://www.asha.org/practice-portal/professional-issues/adult-hearing-screening/

Bakst, S., & Niziolek, C. A. (2019). *Self-correction in L1 and L2 vowel production* [Paper presentation]. Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Vic, Australia.

Banfai, P., Karczag, A., & Lüers, P. (1984). Clinical results: The rehabilitation. *Acta Oto-Laryngologica, 98*(Suppl. 415), 183–194. https://doi.org/10.3109/00016488409121693

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America, 132*(4), 2468–2477. https://doi.org/10.1121/1.4746984

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* [Computer software]. http://www.praat.org

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America, 103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Boucek, M., Ghosh, S., Guenther, F., & Perkell, J. (2008). *A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/* [Paper presentation]. Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France.

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

Cowie, R., & Douglas-Cowie, E. (1992). *Postlingually acquired deafness speech deterioration and the wider consequences* (Vol. 62). De Gruyter Mouton. https://doi.org/10.1515/9783110869125

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America, 70*(1), 45–50. https://doi.org/doi.org/10.1121/1.386580

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

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech, Language, and Hearing Research, 51*(5), 1183–1202. https://doi.org/10.1044/1092-4388(2008/07-0119)

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research, 130*(2), 133–141. https://doi.org/10.1007/s002219900237

Heller Murray, E. S., Lupiani, A. A., Kolin, K. R., Segina, R. K., & Stepp, C. E. (2019). Pitch shifting with the commercially available Eventide Eclipse: Intended and unintended changes to the speech signal. *Journal of Speech, Language, and Hearing Research, 62*(7), 2270–2279. https://doi.org/10.1044/2019_JSLHR-S-18-0408

Heller Murray, E. S., & Stepp, C. E. (2020). Relationships between vocal pitch perception and production: A developmental perspective. *Scientific Reports, 10*(1), Article 3912. https://doi.org/10.1038/s41598-020-60756-2

Houde, J., Gill, J. S., Agnew, Z., Kothare, H., Hickok, G., Parrell, B., Ivry, R. B., & Nagarajan, S. S. (2019). Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration. *The Journal of the Acoustical Society of America, 145*(5), EL372–EL378. https://doi.org/10.1121/1.5100910

Houde, J., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

Houde, J., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5,* 82. https://doi.org/10.3389/fnhum.2011.00082

Huang, X., Chen, X., Yan, N., Jones, J. A., Wang, E. Q., Chen, L., Guo, Z., Li, W., Liu, P., & Liu, H. (2016). The impact of

Parkinson's disease on the cortical mechanisms that support auditory–motor integration for voice control. *Human Brain Mapping, 37*(12), 4248–4261. https://doi.org/10.1002/hbm.23306
Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America, 108*(3, Pt. 1), 1246–1251. https://doi.org/10.1121/1.1288414
Jones, J. A., & Munhall, K. G. (2005). Remapping auditory–motor representations in voice production. *Current Biology, 15*(19), 1768–1772. https://doi.org/10.1016/j.cub.2005.08.063
Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech, 55*(2), 295–308. https://doi.org/10.1177/0023830911417802
Kim, K. S., & Max, L. (2014). Estimating feedforward vs. feedback control of speech production through kinematic analyses of unperturbed articulatory movements. *Frontiers in Human Neuroscience, 8,* 911. https://doi.org/10.3389/fnhum.2014.00911
Kim, K. S., Wang, H., & Max, L. (2020). It's about time: Minimizing hardware and software latencies in speech research with real-time auditory feedback. *Journal of Speech, Language, and Hearing Research, 63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419
Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science, 4*(1), 61–70. https://doi.org/10.1111/1467-7687.00149
Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience, 32*(27), 9351–9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012
Lane, H., Perkell, J., Svirsky, M., & Webster, J. (1991). Changes in speech breathing following cochlear implant in postlingually deafened adults. *Journal of Speech and Hearing Research, 34*(3), 526–533. https://doi.org/10.1044/jshr.3403.526
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. *The Journal of the Acoustical Society of America, 101*(4), 2244–2252. https://doi.org/10.1121/1.418245
Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F_0 control. *Experimental Brain Research, 187*(4), 613–621. https://doi.org/10.1007/s00221-008-1330-z
Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F_0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America, 110*(6), 2845–2848. https://doi.org/10.1121/1.1417527
Leder, S. B., Spitzer, J. B., Milner, P., Flevaris-Phillips, C., Richardson, F., & Kirchner, J. C. (1986). Reacquisition of contrastive stress in an adventitiously deaf speaker using a single-channel cochlear implant. *The Journal of the Acoustical Society of America, 79*(6), 1967–1974. https://doi.org/10.1121/1.393204
Lehiste, I. (1970). *Suprasegmentals*. MIT Press.
Leonard, R. J., & Ringel, R. L. (1979). Vocal shadowing under conditions of normal and altered laryngeal sensation. *Journal of Speech and Hearing Research, 22*(4), 794–817. https://doi.org/10.1044/jshr.2204.794
Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The relation of articulatory and vocal auditory–motor control in typical speakers. *Journal of Speech, Language, and Hearing Research, 63*(11), 3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192
Levitt, H., Stromberg, H., Smith, C., & Gold, T. (1980). The structure of segmental errors in the speech of deaf children. *Journal of Communication Disorders, 13*(6), 419–441. https://doi.org/10.1016/0021-9924(80)90043-x
Li, W., Zhuang, J., Guo, Z., Jones, J. A., Xu, Z., & Liu, H. (2019). Cerebellar contribution to auditory feedback control of speech production: Evidence from patients with spinocerebellar ataxia. *Human Brain Mapping, 40*(16), 4748–4758. https://doi.org/10.1002/hbm.24734
Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex. *The Journal of the Acoustical Society of America, 122*(6), 3671–3677. https://doi.org/10.1121/1.2800254
Liu, H., Zhang, Q., Xu, Y., & Larson, C. R. (2007). Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *The Journal of the Acoustical Society of America, 122*(4), 2405–2412. https://doi.org/10.1121/1.2773955
Liu, P., Chen, Z., Jones, J. A., 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
Maas, E., Mailend, M.-L., & Guenther, F. H. (2015). Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language, and Hearing Research, 58*(2), 185–200. https://doi.org/10.1044/2014_JSLHR-S-13-0300
Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology, 16*(19), 1918–1923. https://doi.org/10.1016/j.cub.2006.07.069
Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience, 33*(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013
Niziolek, C. A., & Kiran, S. (2018). Assessing speech correction abilities with acoustic analyses: Evidence of preserved online correction in persons with aphasia. *International Journal of Speech-Language Pathology, 20*(6), 659–668. https://doi.org/10.1080/17549507.2018.1498920
Niziolek, C. A., Nagarajan, S. S., & Houde, J. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience, 33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013
Niziolek, C. A., Nagarajan, S. S., & Houde, J. (2015). *The contribution of auditory feedback to corrective movements in vowel formant trajectories* [Paper presentation]. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, United Kingdom. https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2015/Papers/ICPHS1004.pdf
Niziolek, C. A., & Parrell, B. (2021). Responses to auditory feedback manipulations in speech may be affected by previous exposure to auditory errors. *Journal of Speech, Language, and Hearing Research, 64*(6S), 2169–2181. https://doi.org/10.1044/2020_JSLHR-20-00263
Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59*(2), 441–449. https://doi.org/10.2307/1130323
Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience, 37*(38), 9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017
Parrell, B., Ivry, R. B., Nagarajan, S. S., & Houde, J. (2021). Intact correction for self-produced vowel formant variability in

individuals with cerebellar ataxia regardless of auditory feedback availability. *Journal of Speech, Language, and Hearing Research, 64*(6S), 2234–2247. https://doi.org/10.1044/2021_JSLHR-20-00270

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology, 15*(9), e1007321. https://doi.org/10.1371/journal.pcbi.1007321

Patel, R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Pail, D., Svec, J. G., & 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. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Patel, R., Niziolek, C., Reilly, K., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research, 54*(4), 1051–1059. https://doi.org/10.1044/1092-4388(2010/10-0162)

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics, 25*(5), 382–407. https://doi.org/10.1016/j.jneuroling.2010.02.011

Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., Vick, J., & Burton, E. (2007). Time course of speech changes in response to unanticipated short-term changes in hearing state. *The Journal of the Acoustical Society of America, 121*(4), 2296–2311. https://doi.org/10.1121/1.2642349

Perkell, J. S., **Lane, H., Svirsky, M., & Webster, J.** (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *The Journal of the Acoustical Society of America, 91*(5), 2961–2978. https://doi.org/10.1121/1.402932

Rathna Kumar, S. B., Azeem, S., Choudhary, A. K., & Prakash, S. G. (2013). The role of auditory and kinaesthetic feedback mechanisms on phonatory stability in children. *Indian Journal of Otolaryngology and Head & Neck Surgery, 65*(Suppl. 3), 562–568. https://doi.org/10.1007/s12070-012-0537-2

Scheerer, N. E., Jacobson, D. S., & Jones, J. A. (2016). Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production. *Neuroscience, 314,* 106–115. https://doi.org/10.1016/j.neuroscience.2015.11.037

Scheerer, N. E., & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience Letters, 529*(2), 128–132. https://doi.org/10.1016/j.neulet.2012.09.012

Scheerer, N. E., Liu, H., & Jones, J. A. (2013). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience, 38*(8), 3189–3200. https://doi.org/10.1111/ejn.12301

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research, 18*(4), 795–811. https://doi.org/10.1044/jshr.1804.795

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *Journal of Speech, Language, and Hearing Research, 63*(7), 2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296

Sorensen, D., Horii, Y., & Leonard, R. (1980). Effects of laryngeal topical anesthesia on voice fundamental frequency perturbation. *Journal of Speech and Hearing Research, 23*(2), 274–283. https://doi.org/10.1044/jshr.2302.274

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature, 423*(6942), 866–869. https://doi.org/10.1038/nature01710

Weerathunge, H. R., Abur, D., Enos, N. M., Brown, K. M., & Stepp, C. E. (2020). Auditory–motor perturbations of voice fundamental frequency: Feedback delay and amplification. *Journal of Speech, Language, and Hearing Research, 63*(9), 2846–2860. https://doi.org/10.1044/2020_JSLHR-19-00407

Wei, K., & Kording, K. (2009). Relevance of error: What drives motor adaptation? *Journal of Neurophysiology, 101*(2), 655–664. https://doi.org/10.1152/jn.90545.2008