Task-Dependent Modulation of Auditory Feedback Control of Vocal Intensity

*,§Allison I. Hilger, *,†Samuel Levant, *Jason H. Kim, *,‡Rosemary A. Lester-Smith, and *Charles Larson,
*Evanston, IL, †Atlanta, GA, ‡Austin, TX, and §Boulder, CO

Summary: Auditory feedback control of fundamental frequency (f_o) is modulated in a task-dependent manner. When voice pitch auditory feedback perturbations are applied in sentence versus sustained-vowel production, larger and faster vocal f_o responses are measured in sentence production. This task-dependency reflects the scaling of auditory targets for pitch for the precision required in each speech task. When the range for the pitch auditory target is scaled down for precision (as in the sentence-production task), a greater degree of mismatch is detected from the feedback perturbation and a larger vocal response is measured. The purpose of this study was to determine whether auditory feedback control of vocal intensity is also modulated in a task-dependent manner similar to the control of vocal pitch. Twenty-five English speakers produced repetitions of a sentence and a sustained vowel while hearing their voice auditory feedback briefly perturbed in loudness (+/- 3 or 6 dB SPL, 200 ms duration). The resulting vocal intensity responses were measured, and response magnitudes were robustly larger in the sentence (mean: 1.96 dB) than vowel production (mean: 0.89 dB). Additionally, response magnitudes increased as a function of perturbation magnitude only in sentence production for downward perturbations but decreased in magnitude by perturbation magnitude for upward perturbations. Peak response latencies were robustly shorter in sentence (mean: 184.94 ms) than in vowel production (mean: 214.92 ms). Overall, these results support the hypothesis that auditory feedback control of pitch and loudness are modulated by task and that both pitch and loudness auditory targets are scaled for the precision required for the speaking task.

Key words: Auditory feedback−Loudness−Perturbation.

INTRODUCTION

Vocal loudness can have a significant impact on the development of voice disorders.[1] Additionally, neurological disorders such as Parkinson's disease can cause reduced control of the vocal intensity, resulting in speech that has variable or reduced loudness and impaired intelligibility.[2,3] Understanding neural control of vocal loudness, the perceptual correlate of vocal intensity, is important for informing voice and speech therapy treatments. The purpose of this study was to better understand the neural control of vocal intensity by measuring auditory feedback control of perceived changes in loudness. Auditory feedback is used during online speech production to correct for perceived errors in speech.[4,5] The identification of errors in speech is related to the scale of the production target; the smaller the production target, the greater the room for error.[6] The goal of this study is to measure how the auditory feedback control system corrects for perceived error in vocal loudness among two speaking tasks to examine task-dependent modulation of feedback control of vocal intensity.

One method for studying auditory feedback control of vocal intensity is to modify auditory feedback of vocal loudness while participants produce sustained vowels and sentences.[7−9] In auditory feedback perturbation studies, participants vocalize into a microphone while their voice is briefly perturbed through headphones in real-time.[4] Most of the research using this technique has measured pitch auditory feedback control. When brief 200 ms pitch perturbations are applied to voice auditory feedback during online speech production, rapid vocal fundamental frequency (f_o) responses are elicited usually in the opposite direction of the perturbation. Because of the rapid nature of these vocal responses, they are assumed to be reflexive in nature as an attempt to correct the perceived error and maintain vocal stability around a target.[4] While most vocal responses occur in the opposite direction of the perturbation (termed *Opposing Responses*), some vocal responses follow the perturbation direction (termed *Following Responses*).[10,11] Some theories posit that response direction is determined by whether the perturbation is perceived as an internal or external error or by the predictability of the perturbation; however, the exact nature of following responses is still unknown.[10] Overall, vocal responses to auditory feedback perturbations are a useful behavioral measure of neural control of voice.

The magnitude of the vocal f_o response has been found to vary as a function of speaking task and perturbation magnitude.[7,9,11,12] For example, larger pitch perturbation magnitudes elicited larger vocal f_o response magnitudes compared to smaller perturbation magnitudes,[11,12] and more complex speaking tasks, such as repeating a sentence or singing a note, elicited larger vocal f_o responses than

Accepted for publication August 2, 2022.
From the *Department of Communication Sciences and Disorders, 2240 Campus Drive, Evanston, IL 60208; †Emory University School of Medicine, 100 Woodruff Circle, Atlanta, GA 30322; ‡Department of Communication Sciences and Disorders, The University of Texas at Austin, 2504A Whitis Ave., Austin, TX 78712; and the §Department of Speech, Language, and Hearing Sciences, The University of Colorado Boulder, 2501 Kittredge Loop Dr, Boulder, CO 80305.
Address correspondence and reprint requests to Charles Larson, Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60201. E-mail: clarson@northwestern.edu
Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■
0892-1997

https://doi.org/10.1016/j.jvoice.2022.08.004

simply holding a sustained vowel.[12,13] We theorize that this task-dependent modulation of auditory feedback control reflects the precision and scale of the vocal control system for achieving the desired production.

Guenther[6] describes how the size of the auditory target is scaled by the speaking task. For example, target regions for speech sounds shrink when speakers are asked to speak more clearly, resulting in more precise articulation.[14] Guenther[6] attributes this variable target region to a strategy employed by the speech motor system called an *economy of effort*: speakers minimize the amount of movement required for production while maintaining intelligibility for the listener.[15,16] Essentially, speakers tune their production in relation to the speaking context and their judgment of the listener's ability to access the information in the speech signal.[16] For example, speakers may adjust their speech differently if they are speaking with a close friend compared to a professional acquaintance, or if they are giving a presentation compared to talking in a more casual setting. By scaling the size of the target region for production, the speech motor system can retain efficiency while maintaining intelligibility.

The economy of effort used by the speech motor system helps explain the task-dependent nature of the auditory feedback control system. Larger vocal f_o responses are observed in tasks such as singing and producing a sentence than for simply holding a vowel sound because the auditory target regions for voice f_o are smaller in the former contexts with less room for error. Singing requires matching pitch to a musical note and speaking requires the production of intonational patterns for pitch. Therefore, both production tasks require a high level of precision and less room for error. Mismatches between the auditory feedback and the auditory target will be greater in these tasks because the auditory target for f_o is smaller. On the contrary, simple sustained-vowel production with no prescribed pitch target requires less precision in f_o so any detected mismatches will be smaller because there is more room for error (i.e., a larger region for the f_o auditory target). Figure 1 displays a schematic adapted from Guenther[6] and Perkell et al.[14] showing that pitch targets for holding a simple sustained vowel compared to saying the word "Nina" in the phrase, "You know Nina?" are scaled for precision. The pitch target for phrase production is represented by a smaller circle than the pitch target for sustained-vowel production, represented by a larger circle. The size of the circle indicates the room for error and precision of the target. Because there is no prescribed pitch to be produced for sustained-vowel production, the pitch target is large with more room for error. For phrase production, the speaker must produce a pitch that is in line with the pitch contour and intonation for the phrase, "You know Nina?" Therefore, the pitch target is more precise with less room for error. Deviation from the pitch contour could change the intonation for the phrase and thus the intended prosodic meaning. When the actual pitch is produced, represented by the red circle, it is outside of the acceptable boundary for both pitch targets, and the speaker will correct for the perceived error. In the schematic, the degree of correction is greater for phrase production (C_p) than for sustained-vowel production (C_s) because there is a greater distance from the actual pitch output and the pitch target for phrase production than sustained-vowel production.

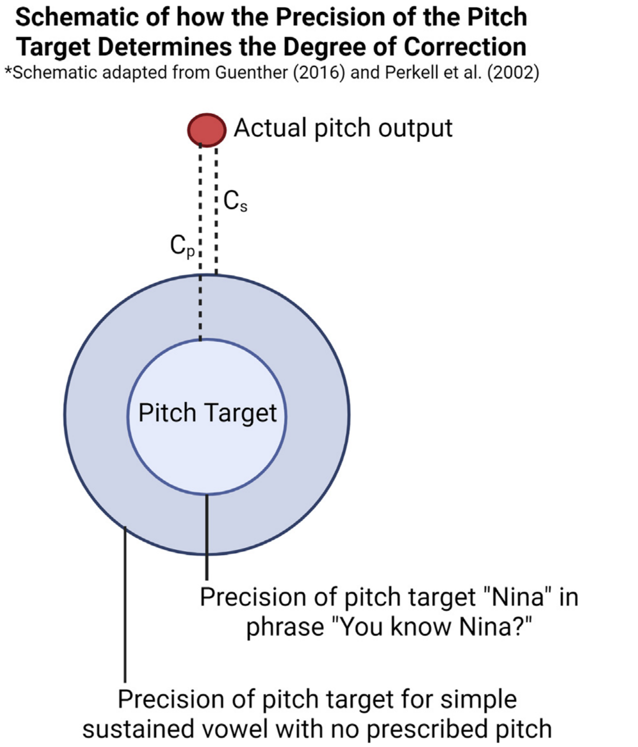
Schematic of how the Precision of the Pitch Target Determines the Degree of Correction
*Schematic adapted from Guenther (2016) and Perkell et al. (2002)
Actual pitch output
C_s
C_p
Pitch Target
Precision of pitch target "Nina" in phrase "You know Nina?"
Precision of pitch target for simple sustained vowel with no prescribed pitch

FIGURE 1. Schematic of how the precision of the pitch target could determine the degree of correction. The pitch target for a simple sustained vowel with no prescribed pitch is represented by the larger circle (dark purple) and the pitch target for saying "Nina" in the phrase, "You know Nina?" is represented by the smaller circle (light purple). The size of these circles represent the precision of the pitch target: the pitch target for phrase production is more precise than for simple sustained-vowel production and is therefore a smaller target with less room for error. The actual pitch output (red circle) is outside the boundaries for both pitch targets, requiring correction. The degree of correction is expected to be greater for phrase production (C_p, longer dashed line), than for sustained-vowel production (Cs, shorter dashed line).

This study examined whether the task-dependency in auditory feedback control of pitch is also observed in control of loudness. A long-standing question in speech modeling research is whether the acoustic correlates of prosody (i.e., pitch, loudness, and timing) are controlled in a similar or different manner to each other.[17,18] If the vocal intensity responses are similarly modulated by speech task and perturbation magnitude as they are for vocal f_o responses, then it is likely that these acoustic features are controlled in a similar manner. This task-dependency has already been observed for perturbation magnitude for vocal intensity responses. In a study using sustained-vowel production, the vocal intensity responses were larger in magnitude for larger loudness perturbations (+/- 3 or 6 dB SPL) than smaller loudness perturbations (+/- 1 dB SPL).[7] The current study expands this work by investigating the control of vocal intensity in sentences compared to sustained-vowel production to determine whether speaking task modulates the magnitude of vocal intensity responses. While rapid loudness perturbations have been applied during phrase production in Mandarin,[9] the novelty of the current study is a within-subject comparison of intensity responses in sentence compared to sustained-vowel production. We hypothesize that auditory feedback control of vocal intensity is modulated by speaking task similar to the pattern observed in auditory feedback control of f_o. If this hypothesis is true, it will demonstrate that auditory targets for vocal intensity are scaled for precision by speaking task.

METHODS

Participants

Twenty-five Northwestern University students between the ages of 18-35 years participated in the study (average age 20 years, standard deviation 3.6 years) (3 female, 22 male). All passed a pure-tone hearing screening according to the American Speech-Language-Hearing Association's guidelines.[19] All participants reported English as their native language and denied a history of speech, language, or neurological disorders. This study was approved by the Northwestern University Institutional Review Board.

Apparatus

After obtaining consent and completing the hearing screening, participants were seated in a double-walled, sound-treated booth (IAC, Model 1201). Participants wore Sennheiser headphones (model HMD 280) and vocalized into an attached AKG microphone (model C420) that was positioned approximately one inch from the corner of the mouth at a 45° angle. Voice output from the microphone was processed, amplified, and perturbed by an Eventide Eclipse Harmonizer and a MOTU Ultralite mk3, which were controlled by MIDI software (Max MSP 5.0, CueMix FX). The auditory feedback was further amplified using an Aphex Headpod 4 Amplifier.

Output from a simplified sound level meter was displayed on a computer screen to aid in the maintenance of a consistent intensity level of approximately 73−75 dB SPL at one inch. In order to mask the participant's bone-conducted feedback, a gain of about 10 dB SPL was applied to the headphone auditory feedback of the participant's voice resulting in an auditory feedback of about 83-85 dB SPL at one inch. All sound and gain calibrations were completed using a Brüel & Kjær sound level meter (type 2250), a set of in-ear microphones (model 4101-A), and a pre-polarized free-field microphone (type 4189) to calibrate the gain between voice output and feedback channels with a 1 kHz sinusoidal pure tone. Providing a sound level meter was used to assist study participants with maintaining a steady vocal intensity both within and across productions.

The perturbed auditory feedback heard by the participants, the true vocal output of the participants, and the timing markers for perturbation onset (using Transistor-Transistor Logic Pulses) were recorded using a multi-channel recording system (AD Instruments, model ML880, PowerLab A/D converter). The PowerLab converter interfaced with LabChart software (AD Instruments, v.7.0) to align the timing markers with participant voice output for later data analysis using Igor Pro 6 (Wavemetrics, Inc., Lake Oswego, OR). The time-aligned markers allow for the determination of the perturbation timing (onset and offset) and direction (upward or downward).

Procedures

Participants followed visual prompts on a computer screen to perform the target vocalizations. The experiment consisted of two sections: a sentence-reading component and a sustained-vowel component. In both components, participants completed four blocks of trials. There were 90 trials/-block for the sentence component and 18 trials/block for the sustained-vowel component. Five perturbations were applied per sustained-vowel trial, and one perturbation was applied per sentence trial. Because the production of a sustained vowel is relatively constant in intensity and f_o, multiple perturbations can be applied in one production. In previous studies, the application of multiple perturbations per vocalization yielded identical results compared to when one perturbation was applied per vocalization.[7,20] Sentence production does not allow for the application of multiple perturbations in one trial because of the fluctuations in intonation across the production of the phrase.

The four blocks of trials consisted of various combinations of pitch and loudness perturbations. For this study, only loudness perturbations were analyzed. Two experimental conditions of the same type (both pitch or both loudness) and of different directions (one upward perturbation and one downward perturbation), as well as an unperturbed control, were included in every block at a 1:1:1 ratio, yielding 30 trials for each condition per block. Block order was partially randomized so that no two loudness perturbation blocks or two pitch perturbation blocks occurred consecutively while still minimizing the block order effects described in.[21] For the loudness blocks, the first block consisted of +6 dB, −3 dB, and control (0 dB) conditions, and the second block consisted +3 dB, -6 dB, and control (0 dB) conditions. Participants were offered rest breaks and water between each block for both the sentence and the vowel components of the study. Overall, there were 16 experimental conditions collected throughout the 8 blocks of the experiment (2 loudness perturbations levels x 2 pitch perturbation levels x 2 perturbation directions x 2 vocal tasks).

The sentence component was completed first to reduce awareness of the perturbations, which were less noticeable in sentence production that sustained vowel production.

The sentence component consisted of four blocks of 90 repetitions of the speech phrase, "You know Nina?" (360 total trials). The intonation of the phrase follows that of a yes/no question in which "you" and "know" are produced with a relatively flat intonation and the final syllable "-na" is produced with a rise in intonation.[22] For each trial, participants first heard an auditory recording of a male speaker producing the sentence presented at about 77 dB SPL in the headphones. Participants were then prompted to repeat the sentence while following visual pacing cues on the screen. The visual pacing cues consisted of four sequential arrows that individually turned bright green at a rate that was consistent with a natural production of the phrase. These visual prompts were provided to assist with the maintenance of a consistent rate of speech across participants and productions and to facilitate correct timing of the perturbations. During each vocalization, a loudness perturbation (+/- 0-, 3-, or 6 dB SPL for 200 ms duration) was randomly presented 640 ms after voice onset so that the perturbation occurred on the "Ni-" syllable of "Nina." The timing of the perturbation was chosen so that it would fall before the rise in intonation on "-na." Participants were given a 1000 ms break before the recorded phrase was played and the next trial started.

After the sentence component was completed, participants performed the sustained-vowel component of the experiment which consisted of four blocks of 18 vocalizations where the participant held an /ɑ/ for approximately 6 s (with 5 perturbations presented per vocalization). Before starting, the task was explained by the experimenter and participants were instructed to keep their voice as steady as possible. Participants were cued by the monitor in the sound booth with a red light and a green light; when the green light was illuminated, participants could begin saying /ɑ/ when they were ready, and participants would continue vocalizing until the red light flashed. Practice trials prior to the experimental trials confirmed that all participants were able to differentiate these two colors and follow the instructions. Each vocalization required producing the /ɑ/ sound for a maximum of 6 s while five perturbations were presented. The first perturbation was presented randomly between 500 ms and 700 ms after voice onset, and the subsequent four perturbations occurred randomly between 700 ms and 900 ms after the prior perturbation. Each perturbation had a duration of 200 ms, and the condition order was determined randomly (in a 1:1:1 ratio, so there were 30 perturbations of every experimental condition in each block).

Analysis

The time-synchronous signals in LabChart were converted to Igor files, and these files were then analyzed using Igor Pro 6 (Wavemetrics, Inc., 2015, Version 6). The perturbed voice signal and the participant's output voice signal were converted to a root-mean-square (RMS) voltage curve for each participant and then converted into decibels. These RMS voltage curves were calculated using the formula which computed RMS for a set of discrete non-overlapping values:

$$\mathrm{RMS}(x) = \sqrt{\frac{1}{N}\sum_{n-25}^{n+25} x^2}$$

where y is the value for each data point and n is the total number of data points. The RMS curves were then segmented in Igor Pro using the timing markers to determine perturbation onset, and segments were defined starting 200 ms before perturbation onset and ending 300 ms after perturbation onset.

To analyze the magnitude and latency of the vocal intensity response, each experimental trial was converted into a difference curve to exclude some of the natural variations in intensity due to intonation. In the target phrase, "You know Nina?" vocal intensity naturally increases during the upward rise in f_o on "Nina" for the yes/no question intonation pattern. The natural variation in intensity needed to be subtracted from the analysis in order to estimate the reflexive intensity response to the perturbation. A custom Igor script calculated an average control curve for each participant and each block. The average control curve included only the trials without perturbations. The average control curve was then subtracted from each experimental trial curve to calculate the difference curve for each trial. Difference curves were also calculated for the sustained vowel task to maintain consistency with the phrase analysis even though we did not suspect any unusual variation in intensity in the vowel production task. A visual analysis was performed on the difference curves to identify curves with measurement or production errors. These curves were identified by their unusual shape (i.e., larger, inconsistent shifts in the intensity contour representing errors in measurement) or extreme values (i.e., changes in intensity greater than 20dB) due to the participant using glottal fry or stopping a vocalization mid-utterance.

After the difference curves were calculated, the pre- and post-perturbation intensity means were compared to determine response direction. More specifically, the pre-perturbation intensity mean of a 200 ms window was compared to the post-perturbation intensity mean of a 300 ms window. If the direction of the response and the direction of the perturbation matched, the trial was labeled as "following;" if they differed, the trial was labeled as "opposing." All trials were categorized opposing vs. following. While there is a possibility that participants occasionally did not respond to the perturbation, it would be difficult to confidently categorize these non-responses for a few reasons. First, the signal is noisy due to constant fluctuations in vocal intensity throughout the production. Taking the average intensity in the post- vs. pre-perturbation window allows for general categorization of whether the response went up or down in intensity. Second, if one could isolate a response in an individual trial, an a priori cut-off value would have to be determined to decide whether variations in intensity constistute a response vs. non-response. With the limited research on loudness-shift responses, it would be difficult to determine what such a cut-off value should be. Therefore, all trials were categorized as opposing vs. following due to constraints of individual-trial noise and difficulty in determining a cutoff value for a response. An event-related averaging technique was then used to reduce the audio signal and allow for extraction of the vocal response.[20] Trials were averaged per participant by perturbation direction (up vs. down), perturbation magnitude (3 vs. 6 dB), and response direction (opposing vs. following), resulting in 200 total potential averaged responses per task (i.e., sentence vs. vowel) (25 participants x 2 perturbation directions x 2 perturbation magnitudes x 2 response directions). Out of the 400 potential averaged responses across the participants and experimental conditions, there were 349 total responses measured (see Table 1). The 51 non-responses included either conditions in which a response was not measured from a participant or responses that were excluded from analysis due to the reasons stated previously (i.e., inconsistent shifts in intensity, extreme values in intensity due to glottal fry).

Response magnitude was calculated by measuring the maximum or minimum peak of the response based on the response direction in a window of 60-300 ms after perturbation onset. This window was chosen because the minimum latency of the vocal response is approximately 60 ms after perturbation-onset, according to the timing of muscular activation and corresponding changes in intensity,[23−25] and to avoid capturing a later volitional response that may occur in the 300-400 ms window.[26] Response latency was then defined as the time-point of the peak response.

Statistical analyses were conducted with R version 4.0.5 (R Core Team, 2021) using RStudio version 1.4.1103 (RStudio Team, 2021). One Bayesian mixed effects model was run using the Stan modeling language[27] and the R package brms.[28] Bayesian modeling was chosen in contrast to frequentist modeling because of the flexible ability to define hierarchical models that include the maximal random effect structure as recommended by Barr et al.[29]. Weakly informative priors were specified for all model parameters. The model used maximal random effect structures, including a random intercept for participants and random slopes allowing the fixed effects to vary by participant.

The model assessed changes in loudness shift response magnitude and latency by task (sentence vs. vowel), perturbation direction (+/- 3 or 6 dB), and response direction (opposing vs. following). For the model predictors, we used regularizing Gaussian priors ($\mu = 0$, $\sigma = 10$) for all variables, signifying that we assumed no effect of the predictors on response magnitude and latency. For the random effects, a half Cauchy distribution was used for the standard deviation ($\mu = 0$, $\sigma = 0.1$) and an LKJ(2) distribution for the correlation. For the residual standard deviation, a half Cauchy distribution was used ($\mu = 0$, $\sigma = 1$).

TABLE 1.
Numbers of Following, Opposing, and Non-Responses for the Phrase and Vowel Task Conditions.

	Sentence	Vowel	Total
Following	79	91	170
Opposing	81	90	179
Total	160	181	349

Four sampling chains with 2,000 iterations were run for each model, with a warm-up period of 1,000 iterations. Ninety-five percent credible intervals (CI's) and probability of direction (*pd*) are reported for each effect. Probability of direction is the probability that a parameter is positive or negative.[30] Given that a value of zero indicates no effect, a higher *pd* value indicates a greater probability that the effect is greater than zero. The 95% CI means that we are 95% certain that the true value lies within the specified interval. The determination of whether there is compelling or robust evidence for an effect is decided whether the 95% interval excludes with zero, and *pd* is greater than 95%.

RESULTS

Of the 349 total responses, there were 179 opposing responses (81 in the sentence task, 98 in the vowel task) and 170 following responses (79 in the sentence task, 91 in the vowel task) (Table 1). A Bayesian generalized linear model was run to determine if the number of opposing or following responses differed by task or perturbation magnitude. Contingent on the data and the model, there were no robust differences in the count of responses for any condition, indicating that speakers opposed and followed the perturbation for all perturbation magnitude and production conditions. See Figure 2 for the averaged responses across all participants and conditions.

Response magnitude

Table 2 displays the 95% credible intervals and median estimates for absolute response magnitude by production task (sentence vs. vowel), perturbation magnitude (+/- 3 or 6 dB), response direction (opposing vs. following), and the interactions among them. Contingent on the data and model, there are two robust main effects for task and perturbation magnitude. First, there is compelling evidence that response magnitude was greater in sentence production than sustained vowel production. Overall, responses were 1.08 dB larger in sentence production (95% CI = [0.80, 1.37]). For perturbation magnitude, response magnitude (averaged across tasks) was larger for -6 dB shifts vs. -3 dB shifts ($\beta = 0.25$ dB, 95% CI = [0.08, 0.43]), -6 dB shifts vs. +6 dB shifts ($\beta = 0.30$ dB, 95% CI = [0.12, 0.47]), and +3 dB shifts vs. +6 dB shifts ($\beta = 0.21$ dB, 95% CI = [0.01, 0.41]) (see Figure 3). Interestingly, larger perturbation magnitudes elicited larger loudness responses only for downward perturbations; the opposite effect was observed for upward perturbations. Note that the robust main effect for

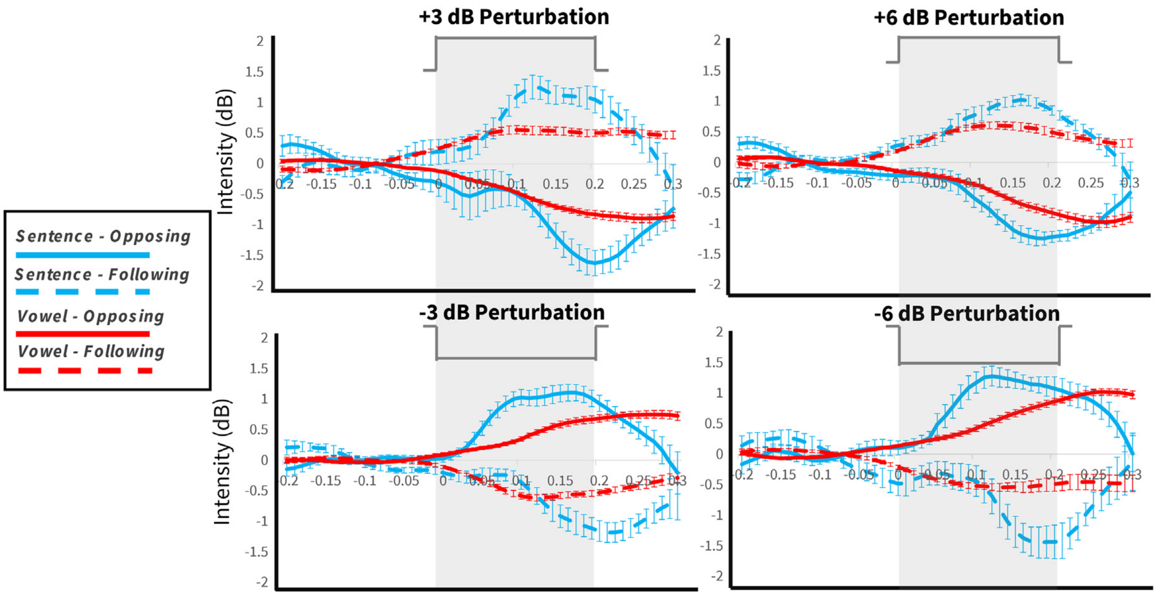

FIGURE 2. Averaged intensity responses by perturbation magnitude and direction, and response direction with intensity (dB) on the Y-axis and time (s) on the x-axis. The averaged intensity responses across participants for opposing responses (solid line) and following responses (dashed line) in sentence production (red line) and sustained vowel production (blue line). In the left-hand column are the responses to 3 dB perturbations and the right-hand column are the responses to the 6 dB perturbations. The top row displays upward perturbations, and the bottom row displays downward perturbations.

perturbation magnitude is averaged across sentence and sustained vowel production and, therefore, overall differences in perturbation magnitude may not be apparent in Figure 2.

There were also robust interactions for perturbation magnitude and task, and perturbation magnitude and response direction. When looking at the interaction between perturbation magnitude and task, the robust differences for response magnitude among perturbation magnitudes were only observed in sentence production. During sentence production, response magnitude was larger for -6 dB shifts vs. -3 dB shifts (β = 0.36 dB, 95% CI = [0.12, 0.460]), -6 dB shifts vs. +6 dB shifts (β = 0.53 dB, 95% CI = [0.27, 0.77]), and +3 dB shifts vs. +6 dB shifts (β = 0.44 dB, 95% CI = [0.14, 0.70]). There were no robust differences by

TABLE 2.
Median Estimate and 95% Credible Interval for the Bayesian Multivariate Mixed Effects Regression Model on the Effect of Task, Perturbation Magnitude, Response Direction, and Their Interactions on Absolute Response Magnitude and Peak Response Latency of the Loudness-Shift Reflex. Probability of Direction (*pd*) Indicates the Probability That the Parameter is Strictly Positive or Negative. Bolded Parameters Indicate Statistically Compelling Evidence for the Effect. Random Effects Estimates are Included for σ^2, Between-Subject Variance (τ_{00}), Intra-Class Coefficient (ICC), Number of Subjects, and Total Number of Observations.

	Response Magnitude			Response Latency		
Predictors	*Estimates*	*CI (95%)*	*pd*	*Estimates*	*CI (95%)*	*pd*
Intercept	1.93	(1.62, 2.23)	100%	0.19	(0.17, 0.20)	100%
Task (Vowel)	**−1.17**	**(−1.49, −0.84)**	**100%**	−0.01	(−0.04, 0.01)	88.48%
Perturbation Magnitude	**−0.06**	**(−0.09, −0.04)**	**100%**	**−0.00**	**(−0.01, −0.00)**	**99.42%**
Response Direction (Opposing)	0.06	(−0.12, 0.23)	75.05%	−0.01	(−0.03, 0.02)	69.75%
Task: Perturbation Magnitude	**0.06**	**(0.02, 0.10)**	**99.95%**	0.00	(−0.00, 0.01)	82.83%
Task: Response Direction	0.19	(−0.04, 0.43)	94.38%	**0.08**	**(0.06, 0.11)**	**100%**
Perturbation Magnitude: Response Direction	**0.08**	**(0.04, 0.11)**	**100%**	**0.01**	**(0.00, 0.01)**	**99.95%**
Task: Perturbation Magnitude: Response Direction	**−0.08**	**(−0.12, −0.02)**	**99.78%**	**−0.01**	**(−0.01, −0.00)**	**98.22%**
Random Effects						
σ^2	0.00					
τ_{00}	0.00					
ICC	0.18					
N $_{subj}$	25					
Observations		349				

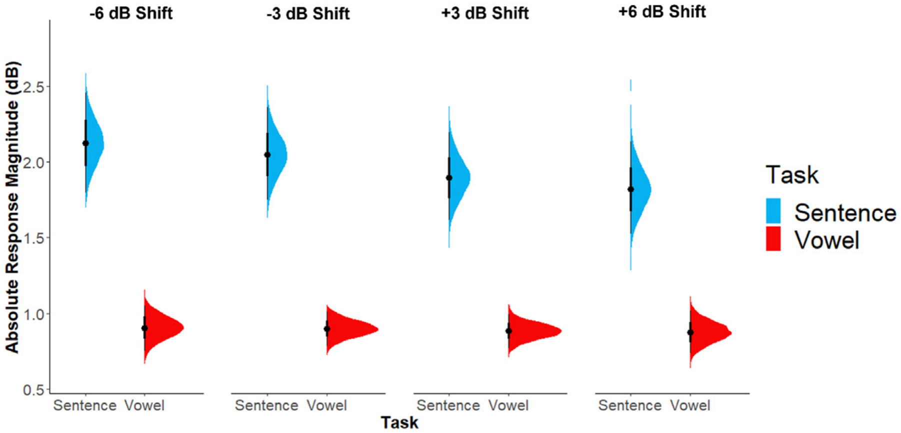

FIGURE 3. Posterior distributions of the absolute response magnitude by perturbation magnitude (−6 dB, −3 dB, +3 dB, +6 dB) and production task (sentence in blue, vowel in red). The full density curves of the model estimates are displayed along with points and whiskers representing the median estimate, 66% credible interval (bolded whisker to the left of the density plot), and 95% credible interval (light whisker).

perturbation magnitude for sustained vowel production. The robust effect of perturbation magnitude on response magnitude was also only robust for following responses. Response magnitude (averaged across tasks) was greater in following responses for −6 dB shifts vs. −3 dB shifts (β = 0.27 dB, 95% CI = [0.03, 0.49]), −6 dB shifts vs. +3 dB shifts (β = 0.24 dB, 95% CI = [0.01, 0.47]), −6 dB shifts vs. +6 dB shifts (β = 0.52 dB, 95% CI = [0.29, 0.74]), −3 dB shifts vs. +6 dB shifts (β = 0.25 dB, 95% CI = [0.02, 0.48]), and +3 dB shifts vs. +6 dB shifts (β = 0.28 dB, 95% CI = [0.04, 0.52]). Note again that because task is not included in this interaction, response magnitude was averaged across sentence and sustained vowel production.

The prevalent pattern across all of these results is that response magnitude was greatest for −6 dB shifts and was lowest for +6 dB shifts, particularly in sentence production and for following responses. Additionally, response magnitude was greater in sentence production than sustained vowel production for all perturbation magnitudes.

Response latency

Table 2 also displays the 95% credible intervals and median estimates for peak response latency by production task (sentence vs. vowel), perturbation magnitude (+/−3 or 6 dB), response direction (opposing vs. following), and the interactions among them. Contingent on the data and model, there is compelling evidence that peak latency was later in vowel production than sentence production for opposing responses (β = −0.07 s, 95% CI = [−0.09, −0.05]) as shown in Figure 4. Within vowel production alone, response latency was shorter for following responses than opposing responses (β = −0.08 s, 95% CI = [−0.10, −0.06]).

Figure 5 displays the complex three-way interaction for peak response latency by task, perturbation magnitude, and response direction. The overall pattern displayed a task effect for peak response latency only for opposing responses: when the loudness response opposed the perturbation direction, the peak response latency was faster in sentence production than in sustained vowel

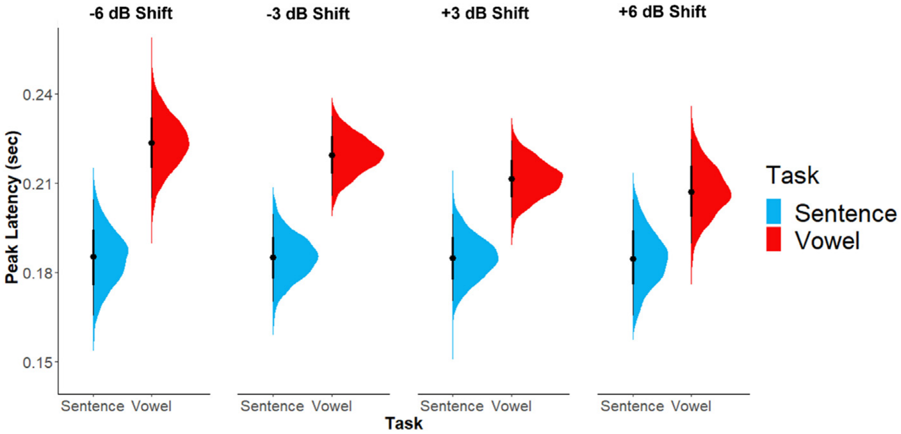

FIGURE 4. Posterior distributions of the peak response latency (s) by perturbation magnitude (−6 dB, −3 dB, +3 dB, +6 dB) and production task (sentence in blue, vowel in red). The full density curves of the model estimates are displayed along with points and whiskers representing the median estimate, 66% credible interval (bolded whisker), and 95% credible interval (light whisker).

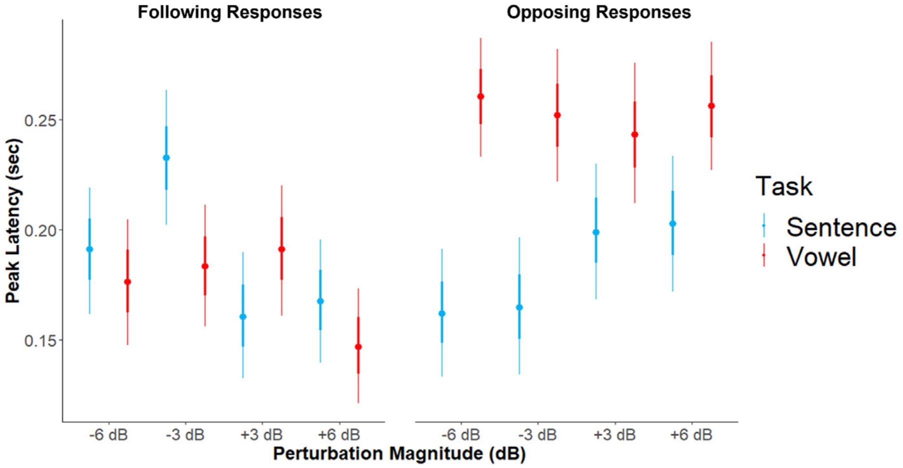

FIGURE 5. Median estimate (point), 66% credible interval (bolded whisker), and 95% credible interval (light whisker) for peak response latency (s) by perturbation magnitude (−6 dB, −3 dB, +3 dB, +6 dB), production task (sentence in blue, vowel in red), and response direction (following vs. opposing).

production. This task effect was not observed for following responses.

DISCUSSION

In this study, auditory feedback control of vocal intensity was found to be modulated by speaking task, similar to control of f_o. When loudness auditory feedback perturbations were applied to ongoing speech, the resulting vocal intensity responses were larger in sentence production than sustained vowel production, similar to findings from[12] using pitch auditory feedback perturbations. These results demonstrate that auditory targets for both intensity and f_o are scaled for the precision required in each task. Error correction was larger in sentence production than sustained-vowel production, supporting the hypothesis that there was a greater degree of mismatch calculated between the scaled auditory target and the perturbation in auditory feedback in the sentence-production task. In other words, it is theorized that when the auditory feedback control system compared the intended output of vocal intensity with the perceived output (i.e., the perturbed loudness feedback), it produced larger corrective responses in sentence production because the intended target had a more specified target range and so a larger response gain was required. As shown in Figure 1, specific auditory targets are required in sentence production to produce the intended prosody for the sentence. Therefore, the auditory target is scaled for precision so that the correct and intended prosodic meaning is conveyed. In contrast, the auditory targets for sustained-vowel production can be highly variable because there are no prescribed pitch or loudness targets to produce. When there is a perceived error in vocal loudness, a greater response gain is needed for correction in sentence production because it is theorized that there is greater distance between the perceived error and the more precise auditory target.

Although modulation of the vocal intensity response by speaking task was observed, there was not a consistent pattern of modulation by perturbation magnitude in which responses were larger for larger perturbations. In sustained-vowel production, vocal intensity responses did not robustly differ in magnitude for 3- or 6 dB perturbations. This finding is consistent with Bauer et al.[7] in which, during sustained-vowel production, there were significant differences in response magnitude for 1 dB perturbations vs. 3- and 6-dB perturbations, but not between 3- and 6-dB perturbations. For the current study, in sentence-production, there were inconsistent patterns of response magnitude by perturbation magnitude. For downward perturbations, intensity responses were robustly larger for −6 dB perturbations than −3 dB perturbations. However, the opposite trend was observed for upward perturbations in which intensity responses were larger for +3 dB perturbations than +6 dB perturbations. These inconsistent patterns of response suggest that auditory feedback control of vocal intensity is not modulated by perturbation magnitude above a 3 dB perturbation. The threshold at 3 dB is specified because there is evidence of modulation by perturbation magnitude for smaller loudness perturbations. In Bauer et al.,[7] vocal intensity responses in sustained-vowel production were significantly smaller in response to 1 dB perturbations than 3 or 6 dB perturbations; additionally, there were no significant differences in response magnitudes for 3 and 6 dB perturbations in this previous study. Taken together, the results from this current study and previous research support the interpretation that perturbation magnitude only modulates vocal intensity responses for smaller loudness perturbations, and that the threshold may be close to 3 dB perturbations. This finding mirrors that from pitch perturbation studies in which beyond a 200-cent perturbation, the error is no longer perceived as self-induced.[31] It is possible that beyond a 3 dB perturbation, the error is similarly no longer perceived as self-induced and a smaller correction gain is applied.

A novel aspect of this study was the isolation of the vocal intensity response to identify response direction (i.e., opposing or following the perturbation direction) by calculating difference curves of the waveforms in a trial-by-trial manner before averaging the trials together. This method allowed for the observation of response direction of the intensity responses regardless of the intonation contour. Without this different curve calculation, it would be difficult to identify opposing or following responses in sentence production. Response direction is identified by comparing the mean intensity in a window after the perturbation to the mean intensity in a window prior to the perturbation and then comparing the direction of the response with the direction of the perturbation. Without subtracting the natural variation in intensity, it would be challenging to identify whether a change in intensity was due to the perturbation or simply the intonation of the sentence. The implementation of a trial-by-trial difference curve analysis is recommended in future perturbation studies to isolate the intensity response and accurately label the response direction.

Focusing on response direction in the current study, there were no robust differences in the number of opposing or following responses by speech task or perturbation magnitude or direction. Response direction has recently been interpreted as a function of whether the speaker perceives the perturbation as an internal error or an external manipulation.[32] There are a few possibilities for the high number of following responses in this experiment. First, it is possible that the constrained nature of the experimental tasks elicited more following responses because participants perceived the perturbations as experimental manipulations. Along this line of thought, it is also possible that the presence of larger perturbations (+/−6 dB) elicited more following responses because they were not always perceived as self-induced errors. Finally, this study categorized response direction differently from previous loudness-perturbation studies by first calculating trial-by-trial difference curves before labeling response direction. Without calculating difference curves, our categorization of response direction could have been confounded by natural variation in intensity. While this final possibility could account for the higher number of following responses in sentence production, given the high variability in intensity due to changes in intonation, it is unlikely to account for the differences in sustained-vowel production compared to previous studies.[7,33] The most likely explanation, therefore, is that participants perceived many of the perturbations as experimental manipulations, resulting in more following responses than usual.

In terms of response magnitude, opposing responses did have robustly larger intensity responses than following responses, an indication that opposing responses may reflect a more corrective mechanism than following responses. Interestingly, there was an interaction between response direction and perturbation direction in which opposing responses to upward perturbations were larger than following responses to upward perturbations. This result may be because speakers perceived upward perturbations as following the intended loudness trajectory, and therefore did not produce as large of a response. For the sentence, “You know Nina?” there is a natural rise in intensity at the end of the sentence that follows the rise in f_o. Given this rise in intensity, it would make sense that speakers would not correct for upward loudness perturbations as much. For response latency, opposing responses were more delayed than following responses, potentially indicating that there is less of a delay to follow a response than to oppose a response.

Peak response latency was also robust for the speech task and response direction. Peak responses in sentence production occurred earlier than in vowel production for opposing responses, indicating that the timing of the vocal responses may be related to the specification and precision of the auditory target. Additionally, there may be a greater demand on the audio-vocal system in sentence production to respond faster than in sustained vowel production because of the timing demands in the production of the sentence. This latter interpretation is supported by a finding in which latency was shorter for vocal f_o responses in sustained vowel tasks when an individual was intending to change f_o.[34] When there are timing demands in production for changing intensity or f_o, the auditory feedback control system may produce more rapid corrective responses, which, in this study, is defined as the time-point of the peak of the response. This interpretation is supported by the Economy of Effort theory described in the introduction.[15,16] The goal of the motor speech system is to retain efficiency while maintaining intelligibility. Therefore, the auditory feedback control system will adjust the speed of response depending on the demands of the production task.

Limitations

Although the results from this study are compelling, there are several limitations that should be addressed. It is possible that having visual feedback of their vocal loudness affected the participants' control of vocal intensity. However, it is highly unlikely that participants were able to adjust the intensity of their voice fast enough to match the movement changes in the sound level meter corresponding to the perturbations.[7] For the purpose of the study, it was more important to assist participants with maintaining a relatively constant vocal intensity than to allow for a potential change in intensity during the perturbations. In natural speech production, speakers tend to reduce their vocal intensity toward the end of an utterance as breath support is decreased.[8]

Another limitation is that the production tasks were highly constrained and may not be representative of typical, everyday speaking patterns. A study on loudness auditory feedback in less constrained speaking tasks is much needed to determine the generalizability of these findings. However, studies with less constrained speaking tasks would require substantial modifications to the current perturbation paradigms. Next, while the analysis in this current study focused on intensity responses to loudness perturbations, the experimental procedure also utilized pitch perturbations in alternating blocks. It is possible that experiencing both pitch and loudness perturbations in a single experimental session could alter the vocal responses to both, although this has not been tested. Lastly, the timing of the loudness perturbation in the sentence is likely to have an impact on the magnitude and timing of the intensity responses. The loudness perturbation was applied on the stressed syllable of the stressed word “Nina” in the sentence, “You know Nina?” Because the perturbation was placed on the stressed word, it is likely that the response is affected by segmental production of the stressed word. Future studies using loudness perturbations would benefit from applying perturbations on different words in a sentence to determine the effect of sentence stress on the intensity response.

CONCLUSION

Overall, the findings of this study demonstrate that auditory feedback control of vocal intensity is modulated by speaking task similarly to auditory feedback control of f_o. Vocal intensity responses to loudness auditory feedback perturbations were larger in magnitude and occurred earlier in sentence-production than sustained-vowel production. This result demonstrates that auditory targets for vocal intensity are scaled by task for precision. Additionally, perturbation magnitude syllables modulated intensity responses to a different extent. The size of the response magnitude did not differ based on the size of the perturbation for 3- and 6 dB perturbations; however, based on prior research, it appears that smaller perturbations of 1 dB elicited a differential response magnitude. It is proposed that there may be a threshold around 3 dB in which the intensity response is scaled by the magnitude of the perturbation. No differences were found for the number of opposing and following responses by task or perturbation condition. However, opposing responses were larger and more delayed than following responses. Overall, these results demonstrate that auditory feedback control of intensity is modulated by speaking task for precision and efficiency.

Acknowledgments

We would like to acknowledge Chun Liang Chan for his expertise and assistance with experiment setup and software development and support.

REFERENCES

1. Bastian RW, Thomas JP. Do talkativeness and vocal loudness correlate with laryngeal pathology? A study of the vocal overdoer/underdoer continuum. *J Voice*. 2016;30:557–562. https://doi.org/10.1016/j.jvoice.2015.06.012.
2. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *J Speech Hear Res*. 1969;12:246–269.
3. Ho AK, Iansek R, Marigliani C, et al. Speech impairment in a large sample of patients with Parkinson’s disease. *Behav Neurol*. 1999;11:131–137.
4. Burnett TA, Freedland MB, Larson CR, et al. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am*. 1998;103:3153–3161.
5. Tourville JA, Guenther FH. The DIVA model: a neural theory of speech acquisition and production. *Lang Cogn Process*. 2011;26:952–981. https://doi.org/10.1080/01690960903498424.
6. Guenther FH. *Neural Control of Speech*. Mit Press; 2016.
7. Bauer JJ, Mittal J, Larson CR, et al. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *J Acoust Soc Am*. 2006;119:2363–2371.
8. Heinks-Maldonado TH, Houde JF. Compensatory responses to brief perturbations of speech amplitude. *Acoust Res Lett Online*. 2005;6:131–137. https://doi.org/10.1121/1.1931747.
9. Liu H, Zhang Q, Xu Y, et al. Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *J Acoust Soc Am*. 2007;122:2405–2412.
10. Behroozmand R, Korzyukov O, Sattler L, et al. Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control. *J Acoust Soc Am*. 2012;132:2468–2477.
11. Liu H, Larson CR. Effects of perturbation magnitude and voice F 0 level on the pitch-shift reflex. *J Acoust Soc Am*. 2007;122:3671–3677. https://doi.org/10.1121/1.2800254.
12. Chen SH, Liu H, Xu Y, et al. Voice F0 responses to pitch-shifted voice feedback during English speech. *J Acoust Soc Am*. 2007;121:1157–1163.
13. Natke U, Donath TM, Kalveram KTh. Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am*. 2003;113:1587–1593. https://doi.org/10.1121/1.1543928.
14. Perkell JS, Zandipour M, Matthies ML, et al. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *J Acoust Soc Am*. 2002;112:1627–1641. https://doi.org/10.1121/1.1506369.
15. Lindblom B. Economy of speech gestures. *The Production of Speech*. New York: Springer; 1983:217–245. https://doi.org/10.1007/978-1-4613-8202-7_10.
16. Lindblom B. Explaining phonetic variation: a sketch of the H&H theory. *Speech Production and Speech Modelling*. Netherlands: Springer; 1990:403–439. https://doi.org/10.1007/978-94-009-2037-8_16.
17. Patel R, Niziolek C, Reilly K, et al. Prosodic adaptations to pitch perturbation in running speech. *J Speech Lang Hear Res*. 2011.
18. Patel R, Reilly KJ, Archibald E, et al. Responses to intensity-shifted auditory feedback during running speech. *J Speech Lang Hear Res*. 2015;58:1687–1694.
19. Association, A. S.-L.-H. (2005). Guidelines for manual pure-tone threshold audiometry.
20. Bauer JJ, Larson CR. Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: improvements in methodology for the pitch-shifting technique. *J Acoust Soc Am*. 2003;114:1048–1054.
21. Scheerer NE, Jones JA. The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *Eur J Neurosci*. 2014;40:3793–3806. https://doi.org/10.1111/ejn.12734.
22. Eady SJ, Cooper WE. Speech intonation and focus location in matched statements and questions. *J Acoust Soc Am*. 1986;80:402–415. https://doi.org/10.1121/1.394091.
23. Kempster GB, Larson CR, Kistler MK. Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *J Voice*. 1988;2:221–229.
24. Larson CR, Kempster GB, Kistler MK. Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. *J Speech Lang Hear Res*. 1987;30:552–558.
25. Perlman AL, Alipour-Haghighi F. Comparative study of the physiological properties of the vocalis and cricothyroid muscles. *Acta Oto-Laryngologica*. 1988;105:372–378.
26. Hain TC, Burnett TA, Kiran S, et al. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res*. 2000;130:133–141. https://doi.org/10.1007/s002219900237.
27. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. *Journal of Statistical Software*. 2017;76.
28. Bürkner P-C. brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*. 2017;80:1–28.

29. Barr DJ, Levy R, Scheepers C, et al. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang*. 2013;68:255–278. https://doi.org/10.1016/J.JML.2012.11.001.
30. Makowski D, Ben-Shachar MS, Chen SHA, et al. Indices of effect existence and significance in the Bayesian framework. *Front Psychol*. 2019;0:2767. https://doi.org/10.3389/FPSYG.2019.02767.
31. Behroozmand R, & Larson CR (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. https://doi.org/10.1186/1471-2202-12-54.
32. Larson CR, Robin DA. Sensory processing: advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neurosci*. 2016;3:22–39. https://doi.org/10.3934/Neuroscience.2016.1.22.
33. Larson CR, Sun J, Hain TC. Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F 0 and amplitude control. *J Acoust Soc Am*. 2007;121:2862–2872.
34. Kim JH, Larson CR. Modulation of auditory-vocal feedback control due to planned changes in voice fo. *J Acoust Soc Am*. 2019;145:1482–1492. https://doi.org/10.1121/1.5094414.