Auditory-Motor Control of Fundamental Frequency in Vocal Vibrato

*[,1]Rosemary A. Lester-Smith, †Jason H. Kim, †[,2]Allison Hilger, ‡Chun-Liang Chan, and †Charles R. Larson,
**Chicago, and* †‡*Evanston, Illinois*

Summary: Purpose. The purpose of this study was to investigate how classically trained singers use their auditory feedback to control fundamental frequency (f_o) during production of vocal vibrato. Two main questions were addressed: (1) Do singers produce *reflexive* f_o *responses* to *sudden perturbation* of the f_o of their auditory feedback during production of vibrato indicative of feedback control? (2) Do singers produce *adaptive* f_o *responses* to *repeated perturbation* of the f_o of their auditory feedback during production of vibrato indicative of feedback and feedforward control? In addition, one methodological question was addressed to determine if adaptive f_o responses were more precisely assessed with or without an auditory cue for f_o during the repeated f_o perturbation paradigm.

Method. Ten classically trained singers produced sustained vowels with vibrato while the f_o and harmonics of their auditory feedback were *suddenly* perturbed by 100 cents to assess *reflexive* control or *repeatedly* perturbed by 100 cents to assess *adaptive* control. Half of the participants completed the repeated perturbation experiment with an auditory cue for f_o, and the other half completed the experiment without an auditory cue for f_o. Acoustical analyses measured changes in mean f_o in response to the auditory feedback perturbations.

Results. On average, participants produced compensatory responses to both sudden and repeated perturbation of the f_o of their auditory feedback. The magnitude of the responses to repeated perturbations was larger than the responses to sudden perturbations. Responses were also larger in the cued, repeated f_o perturbation experiment than in the uncued, repeated f_o perturbation experiment.

Conclusions. These findings indicate that classically-trained singers use both feedforward and feedback mechanisms to control their average f_o during production of vibrato. When compared to prior studies of singers producing a steady voice, the reflexive f_o responses were larger in the current study, which may indicate that the feedback control system is engaged more during production of vibrato.

Key Words: Auditory-motor control−Fundamental frequency−Vocal vibrato.

INTRODUCTION

Typical speakers and experienced singers use their auditory feedback to control their fundamental frequency (f_o) during production of steady, sustained vowels. This phenomenon has been demonstrated in experiments using auditory feedback perturbation paradigms.[1-8] During these experiments, participants vocalized into a microphone and received nearly immediate auditory feedback via earphones. The f_o and harmonics of the auditory feedback was shifted upward or downward, and participants' responses to these f_o perturbations were measured acoustically as changes in f_o. These f_o perturbation studies, also referred to as pitch shift or pitch perturbation studies, demonstrated that speakers and singers primarily produced compensatory f_o responses to perturbations of the f_o of their auditory feedback; that is, their f_o responses were in the opposite direction of the f_o shifts.[1-8] Because the magnitudes of these compensatory responses were found to be larger during singing tasks compared with speaking tasks, it has been suggested that f_o is more precisely controlled in singing.[9]

The f_o perturbation studies that investigated f_o control in experienced singers revealed that musical training, vocal task, and response instructions modulated participants' compensatory responses. Zarate and Zartorre[4,5] found that experienced singers compensated for the f_o perturbations more than non-musicians when both groups were asked to ignore the perturbations, but had equivalent responses when both groups were asked to volitionally compensate for the perturbations. Burnett and Larson[1] demonstrated that trained singers' f_o response magnitudes were greater and response latencies were shorter during production of a steady f_o compared to responses during production of an upward pitch glide. While training, task, and instruction affected singers' compensatory responses, detection of the perturbation did not. That is, Hafke[2] found that trained singers compensated for f_o perturbations, even when the magnitude of the perturbation was below their conscious f_o perturbation detection thresholds. These findings indicated that singing training influences control of f_o, even when f_o

Accepted for publication December 21, 2020.
Declarations of Interest: None.

From the *Department of Physical Medicine & Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; †Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois; and the ‡Department of Linguistics, Northwestern University, Evanston, Illinois.
[1]Present Address: Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin, 2504 A Whitis Avenue, A1100, Austin, TX 78712.
[2]Present Address: Department of Speech, Language, and Hearing Sciences, University of Colorado Boulder, 409 UCB, 2501 Kittredge Loop Drive, Boulder, CO 80309.
Address correspondence and reprint requests to Rosemary A. Lester-Smith, Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin, 2504 A Whitis Avenue, A1100, Austin, TX 78712. E-mail: rosemary.lester-smith@austin.utexas.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■
0892-1997

https://doi.org/10.1016/j.jvoice.2020.12.049

perturbations are below detection thresholds, and that f_o control strategies may differ across vocal tasks.

The aforementioned f_o perturbation studies all assessed singers' *reflexive responses* to sudden perturbations of their auditory feedback. In contrast, Jones and Keough[3] assessed singers' *adaptive responses* to repeated f_o perturbations that were applied and maintained for a block of trials. Responses during the perturbed block were compared to the preceding and following blocks of unperturbed trials. These authors demonstrated that trained singers compensated less during initial exposure to f_o perturbations compared to nonsingers, but that singers maintained their compensatory f_o responses longer after auditory feedback was returned to normal. These results indicated that trained singers may rely less on auditory feedback for control of f_o than nonsingers during initial exposure to f_o perturbations. However, after prolonged exposure to f_o perturbations, singers may learn a new auditory-motor mapping and maintain this new mapping longer than nonsingers.

Singers' reflexive compensatory responses to sudden f_o perturbations and adaptive compensatory responses to sustained f_o perturbations can be accounted for by current models of speech motor control.[10-13] Based on the Directions Into the Velocities of Articulators model,[10,11] a speaker's *feedforward control system* generates an intended voice target or plan, which is conveyed to respiratory, laryngeal, and vocal tract muscles to generate the muscle contractions that result in vocalization. When voice is produced, the output signal is detected by the speaker's auditory system. If a mismatch is detected between the f_o of the intended voice and actual voice output, the *feedback control system* generates an error signal and sends a corrective command to the muscles to immediately adjust the voice output. In addition, the feedback system informs the feedforward system of the error to guide correction of feedforward commands for future error prevention. Therefore, while speakers' reflexive responses to sudden f_o perturbations represent their ability to immediately correct discrepancies between the intended voice and actual voice output using feedback control, speakers' adaptive responses to repeated f_o perturbations represent their ability to gradually correct and prevent discrepancies between the intended voice and actual voice output using both feedback and feedforward control.

To our knowledge, studies of typical auditory-motor control of f_o in singers have all required participants to produce a steady voice and to suppress vocal vibrato. Vocal vibrato is a singing technique that involves desired periodic modulation of the f_o with a modulation rate of 5-7 Hz and a modulation extent of 6%-8%,[14,15] or about a 1 semitone range above and below the average f_o.[16,17] While listeners' perception of the average pitch of vibrato is thought to correspond to the average f_o, based on listeners' perception of modulated pure tones,[18] Sundberg[17] hypothesized that vibrato may mask small inaccuracies in the production of target notes during singing. As such, it is unclear whether singers would need to control the average f_o as precisely during production of vibrato as they would during production of a steady voice. The finding of Burnett and Larson[1] that compensatory responses to f_o perturbations were larger and faster in singers producing a steady voice than in singers producing a pitch glide indicates that the vocal task may influence the detection or correction of f_o errors. Thus, investigating classically trained singers' responses to perturbation of the f_o of their auditory feedback during production of vibrato might clarify how feedforward and feedback mechanisms contribute to expert control of f_o. Furthermore, this investigation could contribute to current models of vocal vibrato.

Titze, Story, Smith, and Long[19] proposed a reflex resonance model of vocal vibrato. In this model, cortical commands to initiate phonation and control average f_o (feedforward commands) are modulated by central oscillators before the commands are sent to the laryngeal muscles. Modulated laryngeal muscle activation then produces modulated vocal fold length and tension. Sensory receptors within the larynx (eg, muscle spindles and joint receptors) detect the vocal fold length modulation and, through a reflexive brainstem network, initiate a motor response that opposes the modulation. This feedback response repeats cyclically and maintains the modulation of vocal fold length and f_o. Using laryngeal electrical stimulation in singers, Titze, Story, Smith, and Long[19] demonstrated that singing training or experience may modulate the scaling and timing of this reflex. It is unknown how this somatosensory reflex interacts with the auditory f_o reflex and if the auditory f_o reflex is similarly affected by singing training, which also influences feedforward control of f_o.[3]

Leydon, Bauer, and Larson[20] demonstrated that, when healthy speakers produced a steady voice and heard auditory feedback with a 1-10 Hz sinusoidally modulated f_o, their voice output became sinusoidally f_o-modulated. The greatest responses were seen when modulation rates were between 4 and 7 Hz, with a peak at 5 Hz (200 ms period). Therefore, auditory feedback might have an important role in controlling f_o in vibrato. Furthermore, auditory and somatosensory mechanisms have been shown to interact in healthy speakers producing steady voice, wherein anesthesia of the laryngeal mucosa increased the magnitude of compensatory responses to perturbation of the f_o of the auditory feedback.[21] Previous studies with healthy speakers producing a steady voice also demonstrated that changes in the f_o of the auditory feedback elicited compensatory contraction of the intrinsic laryngeal muscles that control f_o (ie, cricothyroid and thyroarytenoid muscles).[22] Therefore, further investigation of auditory control of f_o in vibrato is warranted.

The purpose of the present study was to determine how classically trained singers use auditory feedback to control their average f_o during production of vocal vibrato. Two main questions were addressed: (1) Do singers produce *reflexive* f_o *responses* to *sudden perturbation* of the f_o of their auditory feedback during production of vibrato? (2) Do singers produce *adaptive* f_o *responses* to *repeated perturbation* of the f_o of their auditory feedback during production of vibrato? In addition, one methodological question was addressed to determine if adaptive f_o responses were more precisely assessed with or without an auditory cue for f_o during the repeated f_o perturbation paradigm. We hypothesized that classically trained singers would produce both reflexive and adaptive compensatory f_o responses to perturbation of the f_o of their auditory feedback, reflecting both feedback and feedforward control of average f_o during production of vocal vibrato. The alternative hypothesis was that singers would not compensate for perturbation of the f_o of their auditory feedback because they have a wider range of acceptable f_o during production of vibrato. That is, because vibrato involves f_o modulation within a 1 semitone range around the average f_o,[16,17] a detected change in average f_o within this range might not be perceived as an error or corrected. The results of this study could clarify the role of the auditory feedback and feedforward systems in controlling f_o in vocal vibrato. Furthermore, the study findings could have implications for understanding impaired control of f_o in vocal tremor, a neurological voice disorder characterized by involuntary modulation of f_o.[23-25]

METHOD

Participants

Ten classically trained singers (six women and four men) between the ages of 22 and 53 years participated in this study. All participants reported that they were able to speak and read English, were able to follow instructions and pay attention to tasks, had current or past classical voice training, and were able to produce classical vibrato. They denied a current neurological, speech, language, cognitive, or voice disorder; current respiratory disorder affecting speech or voice; and history of surgery on the oral cavity, larynx, pharynx, respiratory system, or central nervous system currently affecting speech or voice. This study was approved by the Northwestern University Institutional Review Board (NU IRB).

Procedure

Prior to data collection, the informed consent process was conducted according to NU IRB guidelines. Participants were informed that the purpose of the study was "to understand how speakers use what they hear to control the voice."

Hearing threshold test

A hearing threshold test was performed with each study volunteer using a Beltone Model 119 or Earscan 3 audiometer with supra-aural headphones. All study volunteers responded to pure tones presented at 25 dB HL at octave intervals between 250 and 4000 Hz in both ears following guidelines for hearing threshold testing in adults.[26]

Interview

Participants were asked a series of questions about their medical, health, and voice history. They were also asked questions about their musical training and experience. Participants reported a range of 4-20 years of singing experience and 4-15 years of singing training. Eight participants reported having choral singing experience. All participants had classical singing training, as well as instrumental experience. Participants' voice types included soprano, mezzo-soprano, countertenor, tenor, baritone, and bass. One participant reported chronic vocal fatigue with prolonged singing. Another participant reported vocal fatigue on the day of the testing due to high voice use prior to the experimental session. This participant also had a history of thyroidectomy. One participant reported that she was previously diagnosed with muscle tension dysphonia and received voice therapy 5 years prior to testing. She denied current or recent symptoms of a voice disorder.

Data collection

Participants were seated in a quiet clinical room for data collection. An AKG C520 head-worn condenser cardioid microphone was positioned 4 cm from the corner of the mouth. The microphone signal was digitized (MOTU UltraLite-mk3) and routed to a laptop computer (Apple MacBook Pro A1278) with CueMix Fx software (MOTU, 2017, Version 1.6 7322). Perturbations of the f_o and harmonics were applied to the digitized signal using a Quadravox harmonizer plug-in (Eventide, 2017, Version 2.3.6), and experimental parameters were controlled by Max software (Cycling '74, 2017, Version 7). The microphone signal was also routed via the MOTU UltraLite-mk3 to a multichannel data acquisition device (ADInstruments PowerLab 8SP ML 785 or 16SP ML 795) connected to a second laptop computer (Apple MacBook Pro A1278) with LabChart software (ADInstruments, 2009, Version 7.0.3). The microphone signal was recorded in LabChart with a sampling frequency of 10 kHz. The perturbed or unperturbed microphone signal was routed via the MOTU UltraLite-mk3 to an earphone amplifier (Aphex HeadPod 4). The amplifier output was routed to insert earphones (Etymotic ER-2) for participants' voice auditory feedback and to the PowerLab for recording of the auditory feedback signal (ie, the perturbed/unperturbed and amplified microphone signal). Deep insertion of the ER-2 ear tips was performed to reduce air conducted feedback of voice output and to reduce the occlusion effect (ie, amplification of bone-conducted feedback of voice; [27]). The earphone amplifier gain was calibrated to be 10 dB SPL louder than the microphone input to mask air conducted feedback of voice output. Calibration was performed using a Brüel & Kjær Type 2250 sound level meter, a 2 cc coupler, and a 1000 Hz pure tone played with a handheld recorder (Olympus VN-541PC) positioned 4 cm from the microphone. Participants received visual cues presented on the laptop computer with Max software.

Repeated fundamental frequency perturbation experiment. *Uncued paradigm.* Participants 1 through 5 completed an uncued repeated f_o perturbation experiment similar to the paradigm used by Jones and Keough[3] and Jones and Munhall[28] to assess adaptive responses.

Participants were instructed to sustain the vowel /ɑ/ for as long as the visual cue "say aaah" appeared on the screen (for 3 seconds) and to take a breath and prepare for the next trial when the visual cue to "breathe" appeared on the screen (for 2-4 seconds randomly jittered). They were instructed to try to find a comfortable note in the first few practice trials that they could produce with vibrato, and to try to produce the same note with vibrato across all trials. They were asked to produce the target intensity represented on a sound level meter on the screen to ensure adequate masking of air-conducted auditory feedback of voice output and to prevent changes in f_o secondary to intensity changes. The target intensity was calibrated to represent a microphone amplitude of 70 dB SPL at 4 cm from the corner of the mouth. Participants were informed that they would hear their voice in their earphones.

Participants completed six practice trials before completing experimental trials in two conditions (ie, control and +100 cent perturbation) with the order pseudorandomized and counterbalanced across participants. In the control condition, participants received unperturbed auditory feedback for 100 trials. The purpose of the control condition was to assess participants' unintended drift in f_o across repeated vowel productions and to allow for normalization of the response magnitudes in the +100 cent perturbation condition to the pattern of drift in the control condition. In the perturbation condition, participants received unperturbed and perturbed auditory feedback across 100 trials in four ordered phases: (1) *baseline,* 25 trials with unperturbed auditory feedback, (2) *ramp*, 20 trials with f_o and harmonics of the auditory feedback gradually increased by 5 cents per trial (ramp trial 1 with 5 cents perturbation; ramp trial 2 with 10 cents perturbation…ramp trial 20 with 100 cents perturbation), (3) *hold*, 30 trials with the f_o and harmonics of the auditory feedback shifted upward by 100 cents, (4) *after-effect*, 25 trials with unperturbed auditory feedback. The maximum perturbation of 100 cents was equivalent to a 1 semitone shift. During perturbed trials, the perturbation was maintained for the full trial period as well as during the inter-trial interval to prevent participants from receiving unperturbed feedback during the ramp and hold phases. Figure 1 represents the magnitude and timing of the perturbation of the auditory feedback relative to the voice output in the hold condition for the maximum +100 cent perturbation.

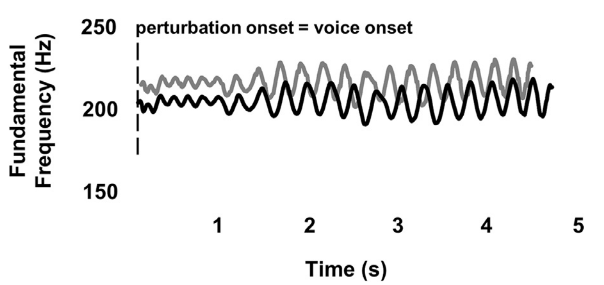

FIGURE 1. Example of a hold trial in the repeated f_o perturbation experiment with the f_o of the microphone signal (black) and the earphone signal (gray) with the +100 cent perturbation applied at voice onset.

Cued paradigm. Participants 6 through 10 completed the cued repeated f_o perturbation experiment. Participants were instructed to sustain the vowel /ɑ/ for a few seconds for three trials to obtain a measurement of their comfortable f_o. The mean f_o of each production was estimated in Praat (Boersma & Weenink, 2017, version 6.0.36). The median f_o of the three trials was entered into Max software, which converted the f_o to the nearest MIDI note to serve as the target note for the experiment. Participants were instructed to listen to the target note before each practice trial and each experimental trial when "listen" appeared on the screen (for 1 second) and to try to match the target note and maintain the same pitch across all trials. For each trial, participants heard the 1 second target note, followed by a visual cue to "breathe," which was presented for 1.5 seconds. Participants were then provided with a visual cue to "say aaah." Presentation of the target note was intended to help maintain the participants' average f_o across the experiment, thereby eliminating the need for a control condition with unperturbed auditory feedback for 100 trials. All other instructions were consistent with the instructions for the uncued repeated f_o perturbation experiment. Participants completed six practice trials before completing the +100 cent perturbation condition as described above.

Sudden fundamental frequency perturbation experiment. Following the repeated f_o perturbation experiment, all participants completed the sudden f_o perturbation experiment. This order was selected to minimize participants' detection of the subtle f_o changes in the repeated f_o perturbation experiment by presenting the more salient f_o changes in the sudden f_o perturbation experiment later in the session. Participants were informed that they would continue to hear their voice in their earphones, but that they may now hear changes in their voice. As in the aforementioned uncued repeated f_o perturbation experiment, they were instructed try to find a comfortable note in the first few practice trials that they could produce with vibrato and then try to produce the same note with vibrato across all trials.

For the practice trials, participants heard upward perturbations of the auditory feedback (+100 cents) in two trials, downward perturbations of the auditory feedback (−100 cents) in two trials, and unperturbed auditory feedback in two trials. Downward f_o perturbations were presented to minimize adaptation to upward perturbations. The order of presentation was randomly determined. Perturbations were applied 1-1.5 seconds after voice onset and were maintained until the end of the trial. Figure 2 represents the magnitude and timing of the perturbation of the auditory feedback relative to the voice output. The intertrial interval was 2-4 seconds and was randomly jittered. Following the practice trials, participants completed 60 randomly ordered experimental trials consisting of 20 trials with +100 cents perturbation, 20 trials with −100 cents perturbation, and 20 trials with unperturbed auditory feedback.

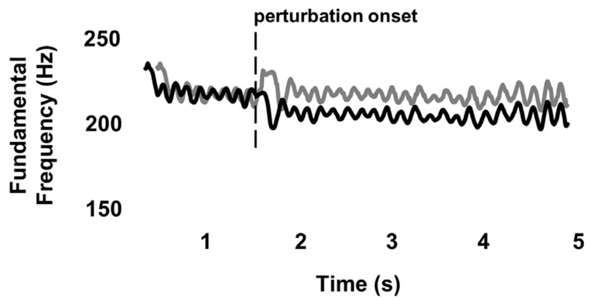

FIGURE 2. Example of a perturbed trial in the sudden f_o perturbation experiment with the f_o of the microphone signal (black) and the earphone signal (gray) with the +100 cent perturbation applied 1.2 seconds after voice onset.

Data analysis

Each trial was visually inspected in Praat (Boersma & Weenink, 2016-2019, versions 6.0.20-6.0.50) using the interface default settings, including the default "pitch floor" of 75 Hz and "pitch ceiling" of 500 Hz. For seven participants, f_o tracking was inconsistent or appeared to be inaccurate for some trials, particularly during brief instances of glottal fry or roughness. To optimize f_o estimation for these participants, the "pitch range" settings were adjusted by raising the pitch floor and lowering the pitch ceiling to be closer to the participant's mean f_o. The individualized pitch range settings were maintained across all experiments and all conditions for each participant. For the first participant, Max software did not trigger timing pulses; therefore, the voice onset was identified manually for each trial performed by this participant. Due to the later finding that Max software occasionally identified voice onset incorrectly (eg, when participants produced a throat clear or tongue click prior to vowel production, the voice onset was detected early; when participants produced a soft voice at onset, the voice onset was detected late), a custom-written Praat script was used to re-identify voice onset for all subsequent participants. Specifically, the Annotate to Text Grid function in Praat was used to identify the correct voice onset when the Max trigger occurred more than 200 ms before voice onset or more than 200 ms after voice onset for multiple trials. For single trials, the mean f_o was obtained manually for 2 seconds after voice onset.

Custom-written Praat scripts were then used to estimate f_o via an autocorrelation method. For the uncued and cued repeated f_o perturbation experiments, the mean f_o was determined for the first 2 seconds of each trial. This analysis window was selected to maintain a consistent window length for both the repeated and sudden f_o perturbation experiments. In addition, although participants were cued to sustain the vowel for 3 seconds, productions were often shorter than 3 seconds due to delays in initiating vowel production after the visual cue to "say aaah" was presented. The mean f_o across baseline trials 6-25 was then calculated. Trials 1-5 were excluded from the baseline mean due to high variability in the f_o as participants acclimated to the task. The response magnitude (in cents) was calculated for each trial using the formula $1200\times \log_2 (f_2/f_1)$, where f_2 was the trial mean f_o and f_1 was the baseline mean f_o. In the uncued paradigm, the +100 cent condition was then normalized to the control condition by subtracting the mean f_o in cents for each control trial from the mean f_o in cents for each corresponding +100 cent condition trial. This normalization to the control condition was performed to account for changes in the participants' f_o across the experiment that were unrelated to the perturbation, which is consistent with previous studies.[29,30] This normalization to the control condition was completed only for the uncued experiment, not the cued experiment. However, both experiments were normalized to their respective baseline phase (ie, trials 6-25).

For the sudden f_o perturbation experiments, the same Praat scripts were used to estimate the mean f_o 1 second prior to the perturbation onset and the mean f_o 1 second following the perturbation onset for the trials with upward perturbation. For the control trials, half of the trials were randomly selected to be analyzed with windows of 0-1 second and 1-2 seconds, and half of the trials were randomly selected to be analyzed with windows of 0.5-1.5 seconds and 1.5-2.5 seconds. These windows were selected to cover the range of analysis used for perturbed trials with the earliest perturbation onset programmed to trigger at 1 second after voice onset and the latest perturbation onset programmed to trigger at 1.5 seconds after voice onset. The response magnitude was calculated for each trial in cents using the formula $1200\times \log_2 (f_2/f_1)$, where f_1 was the pre-perturbation mean f_o and f_2 was the post-perturbation mean f_o for perturbed trials. For the control trials, f_1 was the mean f_o in window 1, and f_2 was the mean f_o in window 2.

Statistical analysis

Data from the repeated f_o perturbation experiment were analyzed using two mixed models with a fixed effect of phase and a random intercept of participant to test for differences in f_o between the baseline and hold phases in the uncued and cued paradigms. The data from the sudden perturbation experiment were analyzed using a third mixed model with a fixed effect of perturbation (ie, +100 or 0 cents) and a random intercept of participant to test for differences in f_o between the perturbed and control trials. The mixed model, also referred to as a random effects model or a hierarchical linear model, accounts for the clustering of data and provides an accurate measure of model and effect variance. The fixed effect in the model is synonymous with the main effect that varies at the level of the participant. The intraclass correlation coefficient (ICC) was estimated by each model, indicating the proportion of variance accounted for by differences between participants. Statistical analyses were performed in R[31] (R Core Team, 2019, v. 3.6.1) using the R package "afex" (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2019, v. 0.25-1).[32]

RESULTS

Repeated fundamental frequency perturbation

Uncued paradigm

On average, participants produced adaptive compensatory responses to the repeated +100 cent perturbation in the ramp and hold phases and maintained some degree of compensation in the after-effect phase (Figure 3). Statistical analyses revealed that the mean f_o during the baseline phase (M = −1.98, SE = 9.18) was significantly higher than the mean f_o during the hold phase (M = −63.57, SE = 9.14), $F(1, 266) = 526.62$, $P = 0.000$, indicating that participants significantly compensated for upward perturbations in the hold phase. Nearly half

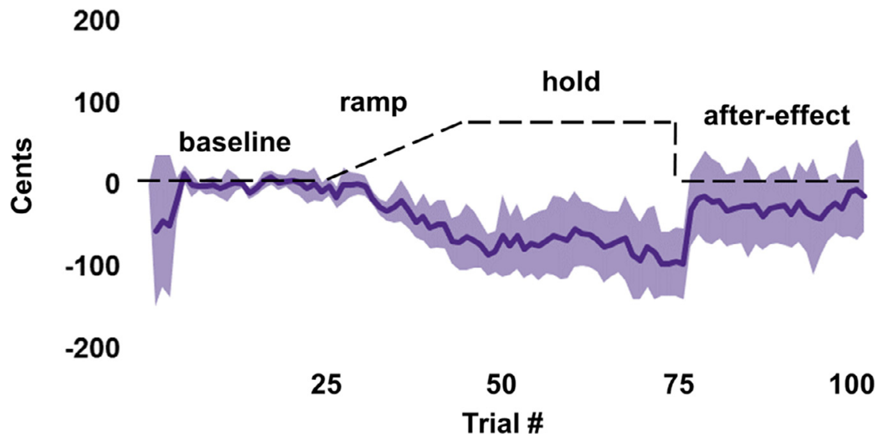

FIGURE 3. Average adaptive responses for Participants 1-5 in the uncued repeated f_o perturbation experiment relative to each participant's baseline mean (trials 6-25) and normalized to each participant's control condition (100 trials with unperturbed auditory feedback; shading = 95% confidence interval; black dashed line = perturbation magnitude).

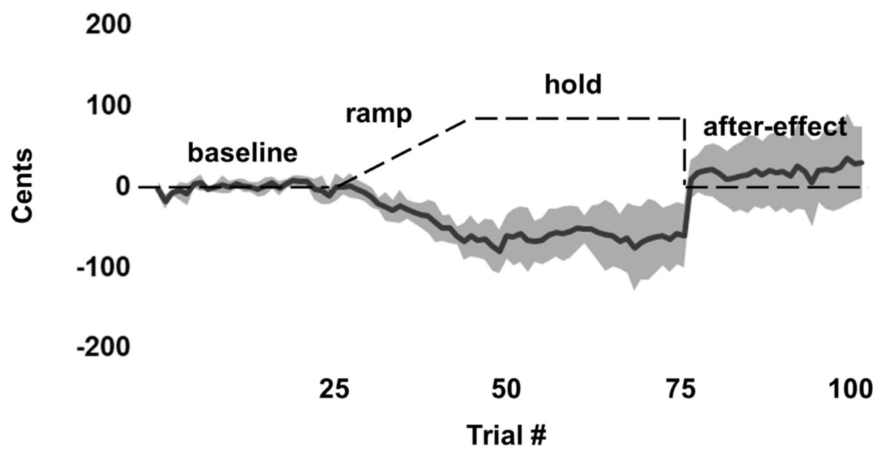

FIGURE 4. Average adaptive responses for Participants 1-5 in the uncued, repeated f_o perturbation experiment relative to each participant's baseline mean without normalization to each participant's control condition.

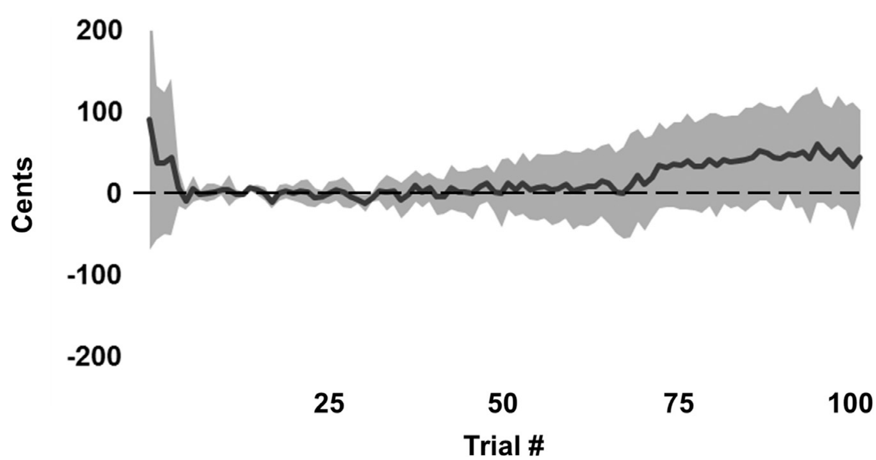

FIGURE 5. Average f_o for Participants 1-5 in the uncued, repeated f_o perturbation experiment control condition relative to each participant's baseline mean.

of the total variability in f_o was related to differences between participants (ICC = 0.45).

Upon inspection of the participants' average response magnitudes before normalization to the control condition (Figure 4), the compensatory responses did not appear to be maintained in the after-effect phase; instead average responses appeared to overshoot the baseline and follow the direction of the perturbation in the previous phase. It was apparent that the after-effect seen in the normalized response was related to the substantial drift in the mean f_o across the 100 trials with unperturbed auditory feedback (Figure 5). The adaptive responses of Participant 3 demonstrated how normalization of the perturbation response to the control condition could alter the estimated responses in both the hold and after-effect phases (Figures 6 and 7) when there was a large upward drift in

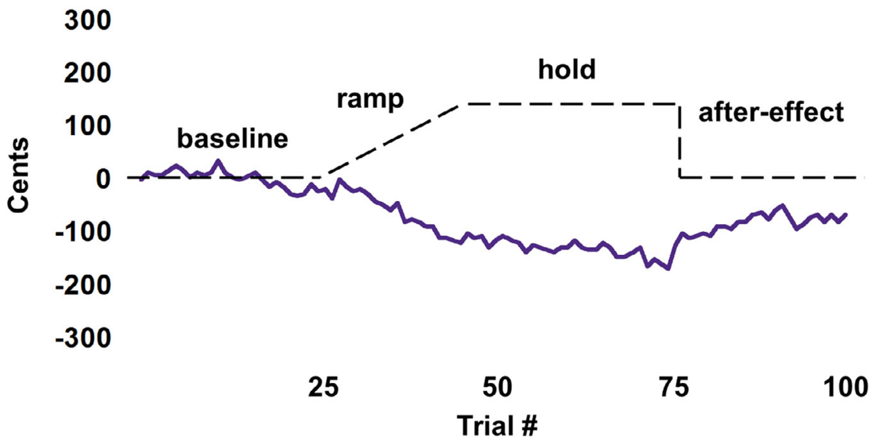

FIGURE 6. The adaptive responses of Participant 3 in the uncued, repeated f_o perturbation experiment relative to the baseline mean and normalized to the control condition.

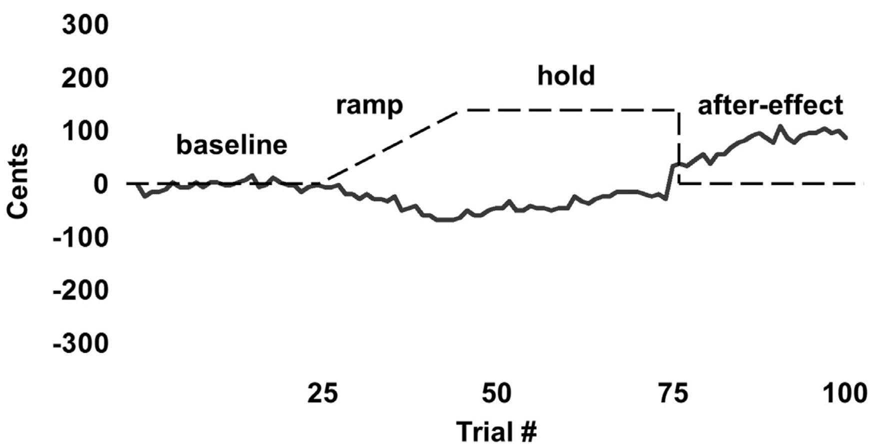

FIGURE 7. The adaptive response of Participant 3 in the uncued, repeated f_o perturbation experiment relative to the baseline mean (not normalized to the control condition).

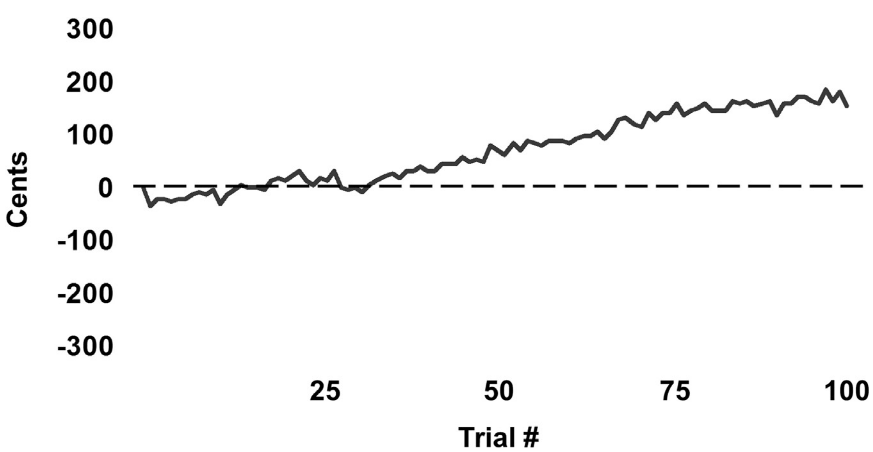

FIGURE 8. The f_o of Participant 3 in the control condition relative to the baseline mean.

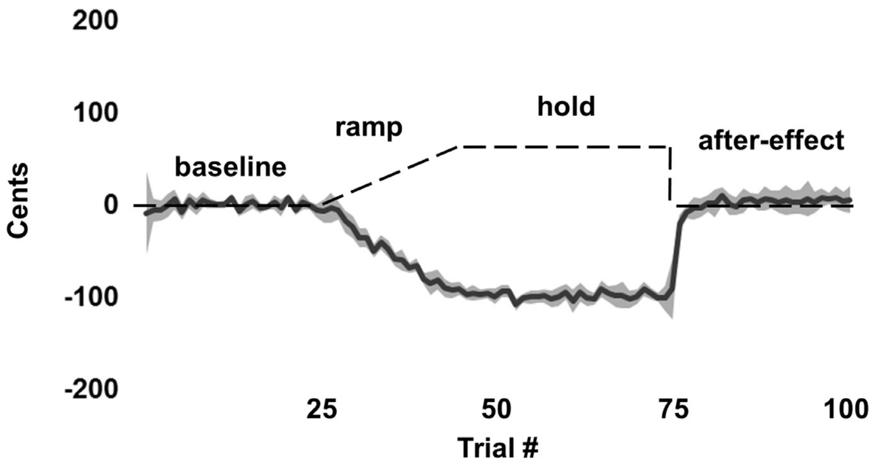

FIGURE 9. Average adaptive responses for Participants 6-10 in the cued, repeated f_o perturbation experiment relative to each participant's baseline mean.

f_o during the control condition (Figure 8). Due to concerns that the normalized responses might not accurately represent participants' compensatory responses to the f_o perturbations, the previously described cued repeated f_o perturbation paradigm was developed and used with subsequent participants.

Cued paradigm

On average, participants produced adaptive compensatory responses to the repeated +100 cent perturbation in the ramp and hold phases of the cued experiment (Figure 9). Statistical analyses revealed that the mean f_o during the baseline phase (M = −0.63, SE = 2.27) was significantly higher than the mean f_o during the hold phase (M = −98.23, SE = 2.23), $F(1, 269) = 5568.4$, $P = 0.000$, indicating that participants significantly compensated for perturbations in the hold phase. About 15% of the total variability in f_o was related to differences between participants, ICC = 0.15.

Sudden fundamental frequency perturbation

On average, participants produced reflexive compensatory responses to sudden +100 cent perturbations of their auditory feedback (Figure 10). In the perturbed condition, 186 responses (93%) were in the compensatory (ie, opposing or negative) direction and 14 responses (7%) were in the following (ie, positive) direction. Statistical analyses revealed that the change in f_o in the perturbed trials (M = −40.93, SE = 3.85) was significantly lower than the change in f_o in the control trials (M = 3.32, SE = 3.85), $F(1, 388.00) = 726.40$, $P = 0.000$, indicating that participants significantly compensated for perturbations. About one-third of the total variability in f_o was related to differences between participants, ICC = 0.33.

DISCUSSION

During production of vibrato, classically trained singers produced adaptive compensatory responses to repeated

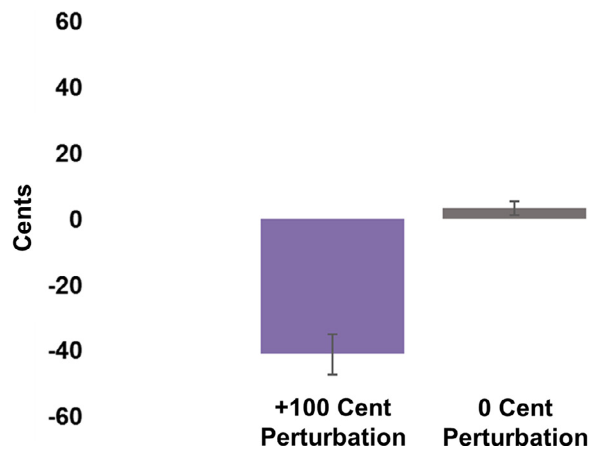

FIGURE 10. Average reflexive responses for Participants 1-10 in the sudden f_o perturbation experiment with +100 cents perturbation (experimental trials) in purple and 0 cents perturbation (control trials) in gray.

perturbation of the f_o of their auditory feedback and reflexive compensatory responses to sudden perturbation of the f_o of their auditory feedback. The average adaptive response magnitudes were −64 cents and −98 cents in the uncued and cued repeated f_o perturbation experiments, and the average reflexive response magnitude was −40 cents in the sudden f_o perturbation experiment. The average f_o in the hold phase was significantly different than the average f_o in the baseline phase for both the uncued and cued repeated perturbation experiments. Similarly, the average f_o in the perturbed trials was significantly different than the average f_o in the control trials for the sudden perturbation experiment. Therefore, 10 singers' data were sufficient for detecting differences between experimental phases and conditions when analyzed using mixed models. Because a substantial proportion of the variance was accounted for by differences between participants, as indicated by the ICC for each model, the mixed model was more appropriate for analyzing these data than a simpler analysis method.

The finding that average adaptive response magnitudes were larger than average reflexive response magnitudes indicates that both the feedback and feedforward systems were involved in controlling average f_o in vibrato. The response magnitudes in the cued and uncued repeated f_o perturbation experiments were similar to the response magnitudes seen in a previous experiment with singers producing a steady voice. Specifically, Jones and Keough[3] found response magnitudes of about 80-100 cents with repeated -100 cent perturbations. Their participants' response magnitudes may have been on the higher end relative to the current findings because they presented an auditory cue for the target note during practice trials. The average response magnitude of -40 cents (40%) in the current sudden f_o perturbation experiment was larger than the response magnitudes seen in previous experiments with singers producing a steady voice. Burnett and Larson,[1] Hafke,[2] and Zarate and Zatorre[4,5] found compensatory responses with magnitudes typically between about 13%-25% of the applied perturbation. The larger reflexive f_o responses in the current study may indicate that the feedback control system is engaged more during production of vibrato than it is during production of a steady voice. Thus, although vibrato might mask listeners' perception of f_o errors in singing as suggested by Sundberg,[17] singers may have a more precise target for the average f_o during production of vibrato than during production of a steady voice. Alternatively, the difference in response magnitudes between these studies and the current study could be attributed to methodological factors including different perturbation magnitudes, directions, and durations; tasks; task instructions; or participant characteristics. Therefore, further research is warranted to compare responses to repeated and sudden f_o perturbations during production of vibrato and steady voice in the same sample of singers.

The current study results indicated that singers use both feedforward and feedback control mechanisms to control average f_o during production of vocal vibrato. Although typical vibrato has an f_o modulation range of 1 semitone,[16,17] singers in the current study corrected for f_o errors of 1 semitone (100 cents). Therefore, despite the fact that a wider range of f_o is expected in vibrato due to the characteristic modulation of f_o around the average f_o, singers still monitor the average f_o of their auditory feedback and adjust for even small errors in production that may be within their vibrato range. The findings that feedback responses were larger during production of vibrato in the current study compared to responses during production of steady voice in prior studies, but that feedforward responses were relatively the same, could imply that singers rely more on feedback control for vibrato than feedforward control. This implication would be consistent with the reflex-resonance model of vocal vibrato, which indicates that the feedforward system is responsible for controlling the average f_o, and reflexive feedback mechanisms are responsible for controlling the extent and rate f_o modulation.[19] It is possible that the feedforward system generates the same motor program for the average f_o during production of vibrato and steady voice, but that the feedback system is engaged more during production of vibrato to control modulation of f_o. Because this study focused on control of average f_o, further research is needed to determine how feedforward and feedback mechanisms are involved in controlling the extent and rate f_o modulation. In addition, future studies should investigate whether the magnitude of the compensatory response is affected by the timing of the perturbation in relation to the phase of modulation within a cycle of vibrato. That is, if an upward perturbation is applied when the f_o is moving in the upward direction during a cycle of vibrato, the magnitude of the compensatory response might be lower than it would be if the f_o were moving in the downward direction.

Finally, as a secondary methodological question, the current study sought to determine if adaptive f_o responses were more precisely assessed with or without an auditory cue for f_o during the repeated f_o perturbation paradigm. The results indicated that the cued repeated f_o perturbation paradigm more precisely assessed feedforward control than the uncued paradigm and eliminated the need for participants to complete 100 control trials with normal auditory feedback. That is, in the cued paradigm, participants' responses were -98 cents as opposed to -64 cents in the uncued experiment. In addition, less variability was seen in both the baseline and the hold phases in the cued gradual f_o perturbation experiment compared to the uncued gradual f_o perturbation experiment. These findings may indicate that an auditory cue strengthens the feedforward program for f_o, leading to enhanced error detection and correction. Therefore, using a cued gradual f_o perturbation experiment to assess feedforward and feedback control may be more sensitive to detecting differences between groups in future studies. Alternatively, these findings may indicate that presenting an auditory cue increases participants' attention to auditory feedback, as previous studies have indicated that greater attention is associated with larger compensatory responses to auditory perturbation.[33,34] Future studies should investigate whether an auditory cue similarly affects singers' feedback control of f_o during sudden perturbation paradigms.

Limitations

To our knowledge, the current study is the first to apply f_o perturbations to modulated voice. Although the perturbed signals retained a sound quality that was realistic and perceived as self-produced sound, there was a processing delay (ie, software and output hardware latency; 35) of about 32 ms. This delay is shorter than delays in previous studies of auditory-motor control of steady voice[29,30]; however, the delay appeared to create a difference in the phase of the f_o modulation in the microphone signal relative to the headphone signal, as can be seen in Figures 1 and 2. This phase difference is not expected to impact the average f_o of the voice output, but it could affect the extent and rate of f_o modulation. Therefore, the current f_o perturbation methods should be modified in future studies to shorten the total feedback loop latency[35] and allow for investigation of feedforward and feedback control of the extent and rate of f_o modulation in vibrato.

In addition to modifying the f_o perturbation methods, analysis methods could be modified for future studies. Specifically, for the sudden f_o perturbation experiment, different acoustical analysis procedures could be used to ensure identical analysis windows for experimental and control trials. In the current study, the f_o perturbation was programmed to trigger 1-1.5 seconds after voice onset. However, when participants produced noise prior to voice onset (eg, a tongue click when they opened their mouth to produce the vowel), the voice onset was detected early causing the perturbation to occur earlier in the vowel production. For the control trials, half were analyzed with windows of 0-1 second and 1-2 seconds after voice onset, and half were analyzed with windows of 0.5-1.5 seconds and 1.5-2.5 seconds after voice onset. Therefore, there may have been more experimental trials than control trials with earlier analysis windows. Although this degree of precision did not appear to be needed for the current study, where control trial responses were substantially different from experimental trial responses, future studies might employ a method of marking or measuring the actual time of perturbation onset relative to voice onset to ensure greater precision in measuring small differences between experimental and control trials.

CONCLUSIONS

This study revealed that classically-trained singers produce adaptive f_o responses to repeated perturbation of the f_o of their auditory feedback and reflexive f_o responses to sudden perturbation of the f_o of their auditory feedback during production of vibrato. When compared to prior studies of singers producing a steady voice, the larger reflexive f_o responses in the current study may indicate that the feedback control system is engaged more during production of vibrato. Furthermore, the study findings indicated that feedforward control of f_o may be strengthened by providing participants with a target note. Further research is warranted to investigate how auditory feedback contributes to control of the extent and rate of f_o modulation in vibrato.

Acknowledgments

We would like to thank Michael Mahometa, PhD, and Erika Hale, MS, in the Department of Statistics and Data Sciences at the University of Texas at Austin for their assistance with the statistical analyses. This research was funded by the National Institute on Disability, Independent Living, and Rehabilitation Research Advanced Rehabilitation Research Training grant 90AR5015 (PI L.R. Cherney) and the National Institute on Deafness and Other Communication Disorders Early Career Research award R21 DC017001 (PI R.A. Lester-Smith).

REFERENCES

1. Burnett TA, Larson CR. Early pitch-shift response is active in both steady and dynamic voice pitch control. *J Acoust Soc Am*. 2002;112:1058–1063.
2. Hafke HZ. Nonconscious control of fundamental voice frequency. *J Acoust Soc Am*. 2008;123:273–278. https://doi.org/10.1121/1.2817357.
3. Jones JA, Keough D. Auditory-motor mapping for pitch control in singers and nonsingers. *Exp Brain Res*. 2008;190:279–287. https://doi.org/10.1007/s00221-008-1473-y.
4. Zarate JM, Zatorre RJ. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Ann N Y Acad Sci*. 2005;1060:404–408. 'doi:'1060/1/404 [pii]10.1196/annals.1360.058.
5. Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*. 2008;40:1871–1887. 'doi:'S1053-8119(08)00059-1 [pii] 10.1016/j.neuroimage.2008.01.026.
6. Zarate JM, Wood S, Zatorre RJ. Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*. 2010;48:607–618.
7. Burnett TA, Freedland MB, Larson CR, et al. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am*. 1998;103:3153–3161.
8. Larson CR, Robin DA. Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neuroscience*. 2016;3:22–39.
9. Natke U, Donath TM, Kalveram KT. Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am*. 2003;113:1587–1593.
10. Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. *Biol Cybern*. 1994;72:43–53.. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=-PubMed&dopt=Citation&list_uids=7880914. Published 1994/01/01.
11. Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*. 2006;96:280–301.
12. Houde JF, Nagarajan SS. Speech production as state feedback control. *Front Hum Neurosci*. 2011;5:82. https://doi.org/10.3389/fnhum.2011.00082.
13. Parrell B, Ramanarayanan V, Nagarajan S, et al. The FACTS model of speech motor control: fusing state estimation and task-based control. *BioRxiv*. 2019 543728.
14. Prame E. Measurements of the vibrato rate of ten singers. *J Acoust Soc Am*. 1994;96:1979–1984.
15. Shipp T, Leanderson R, Sundberg J. Some acoustic characteristics of vocal vibrato. *J Res Sing*. 1980;4:18–25.
16. Seashore CE. The vibrato. *University of Iowa Studies in the Psychology of Music*. 1. Iowa City, IA: University Press; 1932.
17. Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato. In: Dejonckere PH, Hirano M, Sundberg J, eds. *Vibrato*. San Diego, CA: Singular Publishing Group, Inc.; 1995:35–62.
18. Shonle JI, Horan KE. The pitch of vibrato tones. *J Acoust Soc Am*. 1980;67:246–252.
19. Titze IR, Story B, Smith M, et al. A reflex resonance model of vocal vibrato. *J Acoust Soc Am*. 2002;111:2272–2282.
20. Leydon C, Bauer JJ, Larson CR. The role of auditory feedback in sustaining vocal vibrato. *J Acoust Soc Am*. 2003;114:1575–1581.
21. Larson CR, Altman KW, Liu H, et al. Interactions between auditory and somatosensory feedback for voice F 0 control. *Exp Brain Res*. 2008;187:613–621.
22. Liu H, Behroozmand R, Bove M, et al. Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback. *J Acoust Soc Am*. 2011;129:3946–3954.
23. Brown JR, Simonson J. Organic voice tremor: a tremor of phonation. *Neurology*. 1963;13:520–525.
24. Ludlow CL, Bassich CJ, Connor NP, et al. Phonatory characteristics of vocal fold tremor. *J Phon*. 1986;14:509–515.
25. Lester RA, Barkmeier-Kraemer J, Story BH. Physiologic and acoustic patterns of essential vocal tremor. *J Voice*. 2013;27:422–432.

26. American Speech-Language-Hearing Association. *Guidelines for Manual Pure-Tone Threshold Audiometry*. 2005. Rockville, MD.
27. Vasil-Dilaj KA, Cienkowski KM. The influence of receiver size on magnitude of acoustic and perceived measures of occlusion. *Am J Audiol*. 2011;20:61–68.
28. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am*. 2000;108:1246–1251.
29. Abur D, Lester-Smith RA, Daliri A, et al. Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLoS One*. 2018;13: e0191839.
30. Stepp CE, Lester-Smith RA, Abur D, et al. Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *J Speech Lang Hear Res*. 2017;60:1545–1550.
31. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing *[Computer Program]. Version 3.6.1*. Austria: Vienna; 2019.
32. *afex:* Analysis of Factorial Experiments *[Computer Program]*. 2019.
33. Scheerer NE, Tumber AK, Jones JA. Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *J Neurophysiol*. 2016;115:826–832.
34. Tumber AK, Scheerer NE, Jones JA. Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *PLoS One*. 2014;9: e109968.
35. Kim KS, Wang H, Max L. It's about time: minimizing hardware and software latencies in speech research with real-time auditory feedback. *J Speech Lang Hear Res*. 2020;68:2522–2534.