The Effects of Masked and Delayed Auditory Feedback on Fundamental Frequency Modulation in Vocal Vibrato

*,[a]Rosemary A. Lester-Smith, †,[b]Allison Hilger, ‡,[c]Kylie E. Dunne-Platero, †Jason H. Kim, §Chun Liang Chan, and †Charles R. Larson, *Chicago, and †§Evanston, Illinois, and ‡Austin, Texas

Summary: Purpose. Although relatively precise control over the extent and rate of fundamental frequency (f_o) modulation may be needed for optimal production of vibrato, the role of auditory feedback in controlling vibrato is not well understood. Previous studies altered the gain and timing of auditory feedback in singers producing vibrato and showed inconsistent effects on the extent and rate of f_o modulation, which may have been related to small sample sizes or limited analyses. Therefore, the purpose of this study was to further investigate whether the gain or timing of auditory feedback impacts control of vibrato in a larger sample of speakers and with advanced statistical analyses.

Method. Ten classically-trained singers produced sustained vowels with vibrato while their auditory feedback was masked with pink noise or multi-talker babble to reduce the gain of their auditory feedback and while their auditory feedback was delayed by about 200 or 300 milliseconds to alter the timing of their auditory feedback. Acoustical analyses measured changes in the extent and rate of f_o modulation in the masked and delayed trials relative to control trials. Bayesian modeling was used to analyze the effects of noise-masked, babble-masked, and delayed auditory feedback on the extent and rate of f_o modulation.

Results. There was compelling evidence that noise masking increased the extent of f_o modulation, and babble masking increased the variability in the rate of f_o modulation (ie, jitter of f_o modulation). Masked auditory feedback did not affect the average rate of f_o modulation. Delayed auditory feedback did not affect the extent, rate, or jitter of f_o modulation.

Conclusions. The current study demonstrated that reducing the gain of the auditory feedback with noise masking increased the extent of f_o modulation but did not affect the average rate of f_o modulation in classically-trained singers producing vibrato. Reducing the gain of the auditory feedback with babble masking and altering the timing of auditory feedback with imposed delays did not affect the average extent or rate of f_o modulation. However, babble masking increased the jitter of f_o modulation rate, which suggests that modulated auditory feedback may affect the periodicity of f_o modulation from one modulation cycle to the next. These findings clarify the role of auditory feedback in controlling vibrato and may inform the current reflex-resonance models of vibrato.

Key Words: Auditory feedback—Sensorimotor control—Vibrato.

INTRODUCTION

Vocal vibrato is often used in classical singing and involves modulation of the frequency and intensity of voice.[1,2] These acoustical modulations are characterized by: 1) the *extent* or the range of modulation, and 2) the *rate of modulation* or the number of cycles of modulation occurring in one second. The average *extent* of fundamental frequency (f_o) modulation in typical vibrato is 6-8%, or about 1 semitone above and below the average f_o,[2-5] and the average *rate* of f_o modulation in typical vibrato is 4-7 Hz.[1,2,5] Within these ranges, the f_o modulation contributes to a “richness” of tone.[4] However, extents outside the range of ±1 semitone from the average f_o are undesirable and thought to be associated with older age or physical deconditioning.[3,6] Additionally, vibrato rates slower than 5 Hz are perceived as “unacceptably slow” and faster than 8 Hz are perceived as “nervous.”[3] Therefore, relatively precise control over the extent and rate of f_o modulation may be needed for optimal production of vibrato.

Although the exact mechanisms for neural control of vibrato are not well understood, models of speech motor control such as the Directions into the Velocities of the Articulators (DIVA) model[7,8] could have implications for understanding the mechanisms involved in controlling vibrato. In the DIVA model, speech is controlled by *feedforward* and *feedback* systems. The *feedforward* control system creates a motor program for the intended speech that is relayed to muscles of the respiratory, laryngeal, pharyngeal-

Accepted for publication January 30, 2022.

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); the National Institute on Deafness and Other Communication Disorders Early Career Research (Award R21 DC017001) (PI R.A. Lester-Smith); and research funding provided by the Moody College of Communication at The University of Texas at Austin (R.A. Lester-Smith).

From the *Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; †Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois; ‡Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin, Austin, Texas; and the §Department of Linguistics, Northwestern University, Evanston, Illinois.

[a]Current Affiliation: Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin, Austin, TX 78712

[b]Current Affiliation: Department of Speech, Language, and Hearing Sciences, University of Colorado Boulder, 409 UCB, 2501 Kittredge Loop Drive, Boulder, CO 80309

[c]Current Affiliation: Piedmont Healthcare, 710 Center Street, Columbus, GA, 31901

Address correspondence and reprint requests to E-mail: rosemary.lester-smith@austin.utexas.edu

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

https://doi.org/10.1016/j.jvoice.2022.01.029

oral, and velopharyngeal-nasal subsystems for motor execution and speech production. If the sensory system (ie, auditory and somatosensory systems) receives information about the produced speech that does not match the intended speech, the *feedback* system detects this error and responds by immediately sending a corrective command to the muscles. In addition, the feedback system informs the feedforward system of the produced error to update motor programs and prevent future errors.

The feedforward and feedback systems are both implicated in controlling f_o during production of steady, sustained vowels in typical speakers[9-11] and singers.[12-14] Titze, Story, Smith, and Long[6] proposed that the feedforward and feedback systems are also involved in controlling f_o during production of vibrato. In their reflex-resonance model of vocal vibrato, the feedforward system generates a motor program that signals a constant level of muscle activation for the intended f_o. Central oscillators are also activated during production of vibrato and modulate the constant signal. The modulated signal is then transmitted to the laryngeal muscles including the cricothyroid and thyroarytenoid muscles that are involved in controlling f_o. Modulated activation of these muscles produces oscillation of vocal fold length and stiffness, which causes modulation of the f_o. The feedback system is then involved in maintaining the oscillation of vocal fold length and modulation of f_o through a sensorimotor reflex, according to the reflex-resonance model. Specifically, when the cricothyroid muscle contracts, peripheral sensory receptors (ie, muscle spindles) detect the change in vocal fold length. This somatosensory information is immediately relayed to the brainstem, which initiates a counteractive motor response that activates the opposing thyroarytenoid muscle. Contraction of the thyroarytenoid muscle changes vocal fold length, which is detected by the somatosensory system and initiates a subsequent motor response. This sensorimotor reflex alternately produces contraction of the cricothyroid and thyroarytenoid muscles, which maintains modulation of vocal fold length and f_o. The gain of the alternating somatosensory responses determines the level of opposing muscle activation, and the delay in the sensorimotor system determines the timing of the muscle activation. In a typical system, the period of one full cycle—peak to peak cricothyroid activation—is 170 milliseconds or 6 Hz,[6] which is consistent with average rate of f_o modulation in vibrato.[1,2]

The gains and delays of the proposed sensorimotor reflex were tested by Titze, Story, Smith, and Long[6] using computational modeling, as well as mechanical perturbation of the larynx in healthy speakers producing sustained vowels and electrical stimulation of the intrinsic laryngeal muscles in healthy singers producing "vibrato-free" sustained vowels. These studies focused on the role of somatosensory feedback in controlling f_o. However, the speakers and singers received normal voice auditory feedback during the experiments, and Titze, Story, Smith, and Long[6] suggested that auditory feedback might also contribute reflexive control of vibrato. Furthermore, Leydon, Bauer, and Larson[15] demonstrated that modulated auditory feedback induced modulation of f_o in healthy speakers. In their experiment, participants produced sustained vowels while their auditory feedback was sinusoidally modulated at a rate of 1-10 Hz and with an extent of ±25 cents (peak-to-peak modulation of 50 cents). The resultant f_o modulation in the voice output was the greatest when the modulation rates were between 4-7 Hz, suggesting that reflexive auditory-motor control may contribute to the 5-7 Hz f_o modulation associated with vibrato.

More recently, Brajot and Lawrence[16] studied the influence of auditory feedback on voice and speech modulation in typical speakers by delaying the presentation of auditory feedback during sustained vowel production. They found that, with typical auditory feedback, speakers exhibited a 3 Hz modulation of f_o. As the delay in auditory feedback increased from 100-600 milliseconds in 100 milliseconds increments, the extent and rate of f_o modulation linearly increased. Delayed auditory feedback did not significantly affect modulation of the first or second formants in this study. Subsequently, Brajot and Neiman[17] expanded the reflex-resonance model to incorporate an auditory feedback loop.[17] The model was evaluated by altering the auditory feedback gains (ie, the amplitude of the auditory feedback) and delays in a young, healthy speaker and in an individual with intermittent vocal tremor related to multiple sclerosis. This study revealed that the standard deviation of f_o, which represented the f_o modulation extent, increased as the gain increased, whereas the rate of f_o modulation decreased as the feedback delay increased for both participants.

Previous studies also altered the gain and timing of auditory feedback to study auditory-motor control of vocal vibrato. The gain of auditory feedback was altered in two studies using noise masking. Schultz-Coloun and Battmer,[18] as cited by Shipp, Sundberg, and Doherty,[19] masked auditory feedback with "high-level noise" in one singer. They found that the singer's extent of f_o modulation became more variable, although the rate of f_o modulation remained the same. Ward and Burns[20] masked auditory feedback using random low-pass filtered noise in one singer producing vibrato. These authors found that the extent of f_o modulation declined more when the singer was instructed to inhibit vibrato in the presence of masking noise than when they were instructed to inhibit vibrato with normal auditory feedback. However, the presence of masking noise without the instruction to inhibit vibrato did not affect the extent of f_o modulation. Differences in the results of Schultz-Coloun and Battmer[18] and Ward and Burns[20] could have been related to individual differences between the two singers, the procedures used for masking, or the methods used to estimate f_o modulation.

Two other studies altered the timing of auditory feedback by delaying the presentation of the auditory feedback during production of vibrato. Deutsch and Clarkson[21] delayed auditory feedback by 91, 197, 366, and 548 milliseconds in 13 singers and reported that the extent and rate of f_o modulation increased as the delay increased. Similarly, Shipp, Sundberg, and Doherty[19] delayed auditory feedback in three singers and found that f_o modulation rate increased when auditory feedback was delayed by 120, 300, and 500 milliseconds. However, they found that f_o modulation rate did not change when the auditory feedback was delayed by 200 or 400 milliseconds, and they did not find a significant difference in f_o modulation extent or f_o modulation rate variability (ie, jitter of f_o modulation) for any of the delays. The authors suggested that delays of 200 and 400 milliseconds may not have affected f_o modulation extent or rate because the period of the vibrato would have aligned with the duration of the delay. Differences between the findings of Deutsch and Clarkson[21] and Shipp, Sundberg, and Doherty[19] could have been related to the durations of delay, instrumentation used to process and record signals, sample size, analysis approach, or participants' singing training and experience. These conflicting findings from previous studies suggest that additional work using expanded data collection and analysis procedures are needed to clarify of the role of feedback systems in controlling the extent and rate of f_o modulation in vibrato.

The purpose of the present study was to further investigate whether the gain or timing of auditory feedback impacts control of vibrato. Specifically, this study addressed the following question: does masked or delayed auditory feedback affect the extent or rate of f_o modulation in classically-trained singers producing vibrato? Based on the reflex-resonance model,[6] masking somatosensory feedback would be expected to reduce the gain of the somatosensory response, thereby reducing the magnitude of reflexive motor response and the extent of f_o modulation. The recently expanded reflex-resonance model that incorporated an auditory feedback loop[17] would suggest the same. However, we predicted that the somatosensory control mechanisms studied by Titze, Story, Smith, and Long[6] and the auditory-motor control mechanisms studied by Brajot and Neiman[17] during steady, sustained vowel production might differ from auditory-motor control mechanisms during production of vibrato. We hypothesized that singers have an intended extent of f_o modulation that they expect to hear in their auditory feedback, and that masking their auditory feedback would reduce their ability to detect the intended extent of f_o modulation, leading to an increase in the produced extent of f_o modulation. As a secondary question, we investigated whether masking auditory feedback with pink noise or multi-talker babble would have differential effects on the extent of f_o modulation. Because previous studies have indicated that attention to auditory feedback influences auditory-motor control of f_o,[22,23] we hypothesized that the effect would be larger in the babble masking condition than the noise masking condition, as babble would not only mask the singers' air-conducted feedback like noise but might also distract them from attending to their auditory, somatosensory, and bone-conducted feedback. We hypothesized that delaying the auditory feedback would alter timing of the auditory-motor response, thereby changing the timing of the reflexive motor response and the rate of f_o modulation. We further hypothesized based on the findings of Shipp, Sundberg, and Doherty[19] that a delay maintaining an in-phase relationship between the voice output and auditory feedback would not affect vibrato rate; whereas, a delay causing an out-of-phase relationship between the voice output and auditory feedback would reduce the regularity of the rate of f_o modulation. Findings of the present study could improve the understanding of the role of the auditory feedback system in control of vibrato and inform current models of sensorimotor control of voice. Additionally, the methods for altering the gain and timing of auditory feedback developed for the current study could be applied in future studies to investigate essential vocal tremor, a neurogenic voice disorder that has both acoustical and physiological similarities to vibrato[24,25] and may involve impaired speech motor control.

METHOD

Participants

Ten healthy classically-trained singers (six female and four male; ages 22 to 53 years) participated in this study. The same participants completed the f_o perturbation experiments described by Lester-Smith, Kim, Hilger, Chan, and Larson.[26] Participants denied current neurological, speech, language, cognitive and voice disorders. All participants reported at least 4 years of classical singing training and experience. Further details about participant characteristics are reported in Lester-Smith, Kim, Hilger, Chan, and Larson.[26] This study was approved by the Northwestern University Institutional Review Board (NU IRB).

Procedure

All participants completed the informed consent process according to NU IRB guidelines. Participants were informed that the study purpose was "to understand how speakers use what they hear to control their voice."

Hearing threshold test

All participants completed hearing threshold testing and had bilateral hearing thresholds ≤25 dB HL at octave intervals between 250 and 4000 Hz.

Data collection

Procedures and equipment for data collection were similar to those described by Lester-Smith, Kim, Hilger, Chan, and Larson.[26] Data were collected in a quiet clinical room. Each participant was seated in front of a laptop and wore an AKG C520 head-mounted condenser cardioid microphone positioned 4 cm from the corner of the mouth at about a 45 degree angle. The microphone signal was digitized (MOTU UltraLite-mk3) and routed to a multi-channel data acquisition device (ADInstruments PowerLab 8SP ML 785) connected to a laptop computer (Apple MacBook Pro A1278) with LabChart software (ADInstruments, 2009, Version

7.0.3) for signal recording. A 10 kHz sampling rate was used to capture f_o while reducing the file size for each experimental recording. The digitized microphone signal was also routed to a second laptop computer (Apple MacBook Pro A1278) with CueMix Fx software (MOTU, 2017, version 1.6 7322) and Max software (Cycling 74, 2017, Version 7) to control the timing of the experimental trials, visual cues, and auditory feedback. Auditory feedback from Max was sent via the MOTU UltraLite-mk3 to an earphone amplifier (Aphex HeadPod 4) and then to the insert earphones (Etymotic ER-2) as well as the PowerLab for recording of the auditory feedback signal. The ER-2 foam ear tips were inserted deeply to reduce air-conducted and bone-conducted feedback of the produced voice. In addition, auditory feedback was calibrated to be 10 dB SPL louder than the microphone input to mask air-conducted feedback of the produced voice. Levels were calibrated with a Brüel & Kjær Type 2250 sound level meter, 2 cc coupler, and 1000 Hz pure tone presented with a handheld recorder (Olympus VN-541PC) 4 cm from the microphone.

Participants were asked to repeatedly produce the sustained vowel /ɑ/ with vibrato for 3 seconds using a comfortable pitch that they could maintain in each of the three experiments described below. They were instructed to produce the vowel /ɑ/ when a visual cue "aaah" appeared on the laptop screen and to maintain a target loudness based on a sound level meter presented on the same screen. The target intensity was approximately 70 dB SPL at 4 cm from the corner of the mouth. Participants were asked to take a breath and prepare for the next trial when the visual cue "breathe" appeared on the screen for 2-4 seconds (randomly jittered). Participants completed six practice trials before the experimental trials in the three conditions described below. The condition order was pseudorandomized and counterbalanced across participants.

Noise-masked auditory feedback

Participants were informed that they would hear their voice through the earphones for some trials and noise in the earphones for other trials. For the six practice trials, participants were presented with voice auditory feedback for three trials and pink noise to mask their auditory feedback for three trials. These trials were completed in random order. The onset of the noise occurred at the start of the trial prior to voice onset, and a 500 milliseconds ramp in amplitude was applied to prevent a startle response to the 80 dB SPL noise. The noise was presented continuously for 4.5 seconds to ensure that the full 3 seconds vowel production was masked. For the experimental trials, participants received voice auditory feedback for 20 trials (control trials) and pink noise for 20 trials, with the order randomly determined. Example waveforms of the microphone signal and headphone amplifier output signal for a control trial and a noise-masked trial are shown in Figure 1.

Babble-masked auditory feedback

Participants were informed that they would hear their voice in the earphones for some trials and people talking in the earphones for other trials. The practice and experimental trials were identical to those in the noise-masked condition, except that multi-talker babble was presented instead of pink noise. The multi-talker babble was comprised of three male and three female speakers producing sentences, with the signal amplitude fluctuating between about 76 to 84 dB SPL. Example waveforms of the microphone signal and headphone amplifier output signal for a babble-masked trial are shown in Figure 1.

Delayed auditory feedback

Participants were informed that they would hear their voice in the earphones. For the six practice trials, participants were presented with immediate voice auditory feedback for two trials, auditory feedback delayed by about 200 milliseconds for two trials, and auditory feedback delayed by about 300 milliseconds for two trials. The ~200 milliseconds delay was expected to be in phase with the participants' f_o modulation rate, while the ~300 milliseconds delay was expected to be out of phase with the participants' f_o modulation rate based on Shipp, Sundberg, and Doherty.[19] For the experimental trials, participants received voice auditory feedback for 20 trials (control trials), voice auditory feedback delayed by ~200 milliseconds for 20 trials, and voice auditory

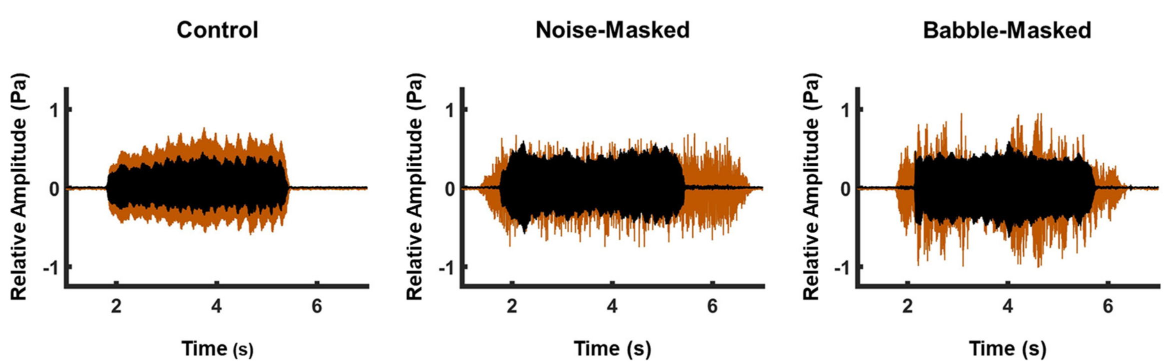

FIGURE 1. Waveforms representing the microphone signal (black) and headphone amplifier output signal (orange) for a control trial (left), noise-masked trial (middle), and babble-masked trial (right) for one participant producing a sustained vowel /ɑ/ with vibrato.

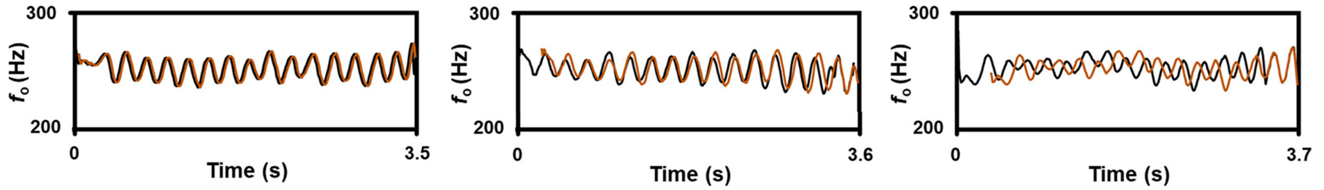

FIGURE 2. Fundamental frequency (f_o) contours representing the microphone signal (black) and headphone amplifier output signal (orange) for a control trial (left), ~200 ms delayed trial (middle), and ~300 ms delayed trial (right) for one participant producing a sustained vowel /ɑ/ with an average vibrato rate of 4.5 Hz in the control trials.

feedback delayed by ~300 milliseconds for 20 trials, with the order randomly determined. Example f_o contours of the microphone signal and headphone amplifier output signal for a control trial, ~200 milliseconds delayed trial, and ~300 milliseconds delayed trial are shown in Figure 2.

Data analysis

Audio recordings were visually inspected in Praat (Boersma & Weenink, 2019-2020, versions 6.0.50-6.1.16) for the accuracy of voice onset identification and f_o tracking for each trial. Timing pulses were generated by Max software during data collection to designate voice onset. When the pulses occurred more than 200 milliseconds before or after the onset of the f_o trace, voice onset was re-identified using a custom-written Praat script that created a text grid and marked periods of sound and silence. The onset of each sound period was then used to designate voice onset. When f_o tracking appeared to be inconsistent or inaccurate (eg, during instances of glottal fry or aperiodicity), the default pitch range of 75-500 Hz was adjusted. The pitch range was incrementally adjusted around the participant's mean f_o until the f_o trace appeared to be consistent with the f_o represented by the narrowband spectrogram. When the default settings were changed for a participant in one experimental condition, the same pitch range was used for all three conditions for a given participant to ensure consistency in the analyses.

Estimates of the f_o for the first 2 seconds of each trial were then obtained using custom-written Praat scripts, which created a pitch object, smoothed the pitch object with a 10 Hz bandwidth, and identified the minimum and maximum f_o peak values and times. Extent was calculated for each modulation cycle in the 2 seconds window using the formula: $(f_{max} - f_{min}) / (f_{max} + f_{min}) \times 100$. The average f_o modulation extent was then determined for each trial. The cycle period was calculated as the time difference between the peak of one cycle and the peak of the preceding cycle. The average f_o modulation rate was determined for each trial using the formula $1/T$, where T was the average cycle period for the trial. Jitter of f_o modulation was also calculated to estimate variability in the rate of f_o modulation from one cycle of modulation to the next, based on Shipp, Sundberg, and Doherty.[19] The difference between the period of each cycle of f_o modulation and the preceding cycle of modulation was determined. An average period difference was then calculated for each trial and converted to a percentage change. In order to determine the actual delay induced by Max software during the delayed auditory feedback experiments, cross-correlation analyses of the recorded microphone signal and the recorded headphone amplifier output signal were completed in Matlab (The Mathworks, Inc., 2018, version R2018a).

Statistical analysis

Statistical analyses were conducted with R[27], version 4.0.5, using RStudio[28], version 1.4.1103. Measures of f_o modulation extent, rate, and jitter were submitted to Bayesian hierarchical models using Stan modeling language[29] and the R package brms.[30] A detailed description of Bayesian statistics is beyond the focus for this paper; however, please refer to Nalborczyk, Batailler, Lúvenbruck, Vilain, and Bürkner[31] for a tutorial on applying Bayesian statistics to speech acoustics research. Bayesian modeling was chosen over frequentist modeling because of the flexibility in defining hierarchical models that include maximal random effect structure as recommended by Barr, Levy, Scheepers, and Tily.[32]

To assess the effects of masked and delayed auditory feedback on f_o modulation extent, rate, and jitter, we fit two separate Bayesian hierarchical regression models. The first model was fit to f_o modulation extent, rate, and jitter predicted by masked auditory feedback (three conditions: control trials, noise-masked trials, and babble-masked trials). The second model was fit to the same dependent variables (f_o modulation extent, rate, and jitter) predicted by delayed auditory feedback (three conditions: control trials, ~200 milliseconds delayed trials, and ~300 milliseconds delayed trials). Both models included maximal random-effect structures with a random intercept for participants and random slopes that allowed the effects of masked and delayed auditory feedback to vary by participant.

Weakly informative priors were specified for all model parameters. For the model predictors (masked and delayed auditory feedback), we used regularizing gaussian priors ($\mu = 0$, $\sigma = 2$), signifying that we assumed no effect of masked or delayed auditory feedback on the dependent variables. For the random effects, a half Cauchy distribution ($\mu = 0$, $\sigma = 0.1$) was used for the standard deviation and an LKJ (2) distribution for the correlation. For the residual standard deviation, a half Cauchy distribution was used (μ = 0, σ = 1). Four sampling chains with 2,000 iterations were run for each model, with a warm-up period of 1,000 iterations. The 95% credible intervals (CI) and the posterior probability that the masked or delayed auditory feedback coefficient was smaller or larger than zero Pr (β> or < 0) are reported below. The 95% CI indicated 95% certainty that the true value lay within the specified interval. When the 95% interval did not overlap with zero and when Pr (β> or < 0) was close to zero, we concluded that there was compelling evidence for an effect.

RESULTS

The results of the f_o modulation extent, rate, and jitter analyses for the masked and delayed auditory feedback experiments are reported for each participant in Tables 1 and 2, respectively. Average results and statistical analyses are presented below.

Masked auditory feedback experiments

The average f_o modulation extent was 2.1% (SD = 1.0) in the control trials, 2.4% (SD = 0.9) in the noise-masked trials, and 2.1% (SD = 1.0) in the babble-masked trials. The

TABLE 1.
Average Results of the Acoustical Analyses for Each Participant in the Masked Auditory Feedback Experiments, Presented as Mean (Standard Deviation)

Participant	Experiment	f_o Mod Extent (%)		f_o Mod Rate (Hz)		f_o Mod Jitter (%)	
		Control	Experimental	Control	Experimental	Control	Experimental
P1	Noise	4.0 (0.5)	4.1 (0.6)	5.0 (0.2)	5.5 (0.2)	2.1 (1.1)	1.8 (0.6)
	Babble	3.4 (0.3)	3.5 (0.6)	5.1 (0.4)	5.5 (0.2)	2.6 (1.2)	2.5 (1.7)
P2	Noise	1.3 (0.4)	1.5 (0.4)	4.3 (0.4)	4.3 (0.2)	3.7 (1.3)	4.5 (2.1)
	Babble	1.2 (0.2)	0.9 (0.3)	4.3 (0.4)	4.3 (0.4)	4.5 (3.2)	6.4 (4.0)
P3	Noise	1.4 (0.2)	1.8 (0.3)	5.2 (0.1)	5.1 (0.1)	1.5 (0.5)	1.6 (1.1)
	Babble	1.2 (0.2)	1.5 (0.3)	5.2 (0.3)	5.3 (0.2)	3.1 (1.1)	2.7 (2.5)
P4	Noise	1.5 (0.5)	2.0 (0.7)	6.0 (0.5)	5.3 (0.4)	2.9 (1.0)	3.5 (1.1)
	Babble	2.1 (0.5)	2.5 (0.6)	5.7 (0.4)	5.2 (0.3)	2.8 (0.9)	2.8 (1.1)
P5	Noise	3.6 (0.6)	3.7 (0.5)	4.6 (0.2)	4.6 (0.2)	1.0 (0.3)	1.4 (0.7)
	Babble	3.4 (0.7)	3.6 (0.5)	4.5 (0.2)	4.5 (0.1)	1.4 (0.6)	1.1 (0.3)
P6	Noise	1.2 (0.2)	1.3 (0.3)	3.6 (0.2)	3.7 (0.2)	3.1 (1.4)	3.9 (2.8)
	Babble	1.1 (0.3)	1.3 (0.3)	3.7 (0.2)	3.8 (0.4)	4.0 (3.3)	4.9 (2.3)
P7	Noise	2.9 (0.3)	2.8 (0.5)	4.6 (0.1)	4.7 (0.1)	1.1 (0.4)	1.7 (0.6)
	Babble	2.7 (0.3)	2.7 (0.3)	4.5 (0.2)	4.7 (0.1)	1.3 (0.5)	1.8 (0.9)
P8	Noise	2.6 (0.4)	2.1 (0.2)	4.3 (0.3)	4.3 (0.2)	2.6 (0.9)	3.0 (1.2)
	Babble	2.6 (0.3)	2.1 (0.3)	4.5 (0.3)	4.5 (0.3)	2.3 (1.2)	3.4 (1.8)
P9	Noise	1.4 (0.4)	1.8 (0.4)	5.3 (0.7)	5.0 (0.4)	5.3 (1.8)	3.3 (1.2)
	Babble	0.9 (0.4)	0.9 (0.3)	5.2 (0.7)	5.3 (0.7)	6.9 (2.4)	7.6 (4.6)
P10	Noise	2.1 (0.1)	2.6 (0.2)	4.6 (0.2)	4.7 (0.1)	1.7 (0.6)	1.6 (0.8)
	Babble	1.9 (0. 3)	2.3 (0.3)	4.7 (0.2)	4.7 (0.1)	1.8 (1.7)	1.6 (0.6)

TABLE 2.
Average Results of the Acoustical Analyses for Each Participant in the Delayed Auditory Feedback Experiment, Presented as Mean (Standard Deviation)

Participant	f_o Mod Extent (%)			f_o Mod Rate (Hz)			f_o Mod Jitter (%)		
	Control	~200 ms	~300 ms	Control	~200 ms	~300 ms	Control	~200 ms	~300 ms
P1	4.1 (0.5)	3.6 (0.6)	3.6 (0.7)	5.1 (0.3)	5.1 (0.3)	5.2 (0.6)	2.1 (1.2)	3.3 (2.0)	2.2 (0.9)
P2	0.7 (0.4)	0.7 (0.3)	0.8 (0.4)	4.3 (0.5)	4.6 (0.8)	4.5 (0.5)	8.6 (5.1)	7.7 (6.8)	8.5 (5.1)
P3	1.3 (0.3)	1.3 (0.3)	1.4 (0.3)	5.2 (0.2)	5.3 (0.2)	5.3 (0.1)	2.2 (1.0)	1.7 (0.8)	1.8 (0.6)
P4	1.4 (0.4)	1.5 (0.3)	1.5 (0.5)	6.0 (0.6)	6.1 (0.4)	5.9 (0.6)	2.9 (0.9)	2.2 (0.8)	2.6 (1.8)
P5	3.3 (0.5)	3.8 (0.5)	2.9 (0.4)	4.5 (0.2)	4.4 (0.1)	4.6 (0.3)	1.3 (0.5)	1.0 (0.4)	1.1 (0.4)
P6	0.9 (0.2)	1.0 (0.2)	1.0 (0.2)	3.8 (0.5)	4.0 (0.3)	3.6 (0.4)	4.2 (2.2)	4.2 (3.4)	6.2 (2.7)
P7	2.8 (0.3)	2.7 (0.3)	2.1 (0.3)	4.5 (0.1)	4.6 (0.1)	5.0 (0.2)	1.1 (0.5)	1.5 (0.6)	1.7 (0.8)
P8	2.6 (0.4)	2.5 (0.5)	2.2 (0.4)	4.5 (0.3)	4.4 (0.4)	4.2 (0.4)	2.7 (1.4)	3.0 (1.7)	4.2 (1.7)
P9	0.7 (0.1)	0.7 (0.2)	0.6 (0.2)	5.2 (0.6)	5.3 (0.7)	5.3 (0.7)	8.2 (2.8)	6.5 (2.1)	8.0 (3.6)
P10	1.8 (0.2)	1.7 (0.2)	1.9 (0.3)	4.6 (0.2)	4.6 (0.2)	4.8 (0.1)	1.7 (0.7)	1.6 (0.7)	1.4 (0.5)

average f_o modulation rate was 4.7 Hz (SD = 0.6) in the control trials, 4.7 (SD = 0.5) Hz in the noise-masked trials, and 4.8 Hz (SD = 0.5) in the babble-masked trials. The average f_o modulation jitter was 2.8% (SD = 1.5) in the control trials, 2.6% (SD = 1.1) in the noise-masked trials, and 3.5% (SD = 2.2) in the babble-masked trials.

The 95% credible intervals and mean estimates for f_o modulation extent, rate, and jitter by condition are shown in Figure 3 and Table 3. Contingent on the data and model, there was compelling evidence that, compared with control trials, noise masking increased f_o modulation extent (β = 0.28, 95% CI = [0.00, 0.56]; Pr ($\beta > 0$) = 0.02) (Figure 4), and babble masking increased f_o modulation jitter (β = 0.36, 95% CI = [0.01, 0.77]; Pr ($\beta > 0$) = 0.02) (Figure 5). There was also evidence, though not compelling, that noise masking increased f_o modulation extent more than babble masking (β = 0.30, 95% CI = [−0.05, 0.68]; Pr ($\beta < 0$) = 0.05) (Figure 6), and babble masking increased f_o modulation jitter more than noise masking (β = 0.39, 95% CI = [−0.15, 0.89]; Pr ($\beta < 0$) = 0.06) (Figure 7).

Delayed auditory feedback experiment

The average f_o modulation extent was 2.0% (SD = 1.2) in the control trials, 2.0% (SD = 1.1) in the ~200 milliseconds delay trials, and 1.8% (SD = 1.0) in the ~300 milliseconds delay trials. The average f_o modulation rate was 4.8 Hz (SD = 0.6) in the control trials, 4.8 Hz (SD = 0.6) in the ~200 milliseconds delay trials, and 4.8 Hz (SD = 0.6) in the ~300 milliseconds delay trials. The average f_o modulation jitter was 3.5% (SD = 2.7) in the control trials, 3.2% (SD = 2.2) in the ~200 milliseconds delay trials, and 3.8% (SD = 2.8) in ~300 milliseconds delay trials.

The cross-correlation analyses revealed that the measured timing difference between the microphone signal and the recorded headphone amplifier output signal was 24 milliseconds for the control trials and 240 and 340 milliseconds for the delayed trials for nine of the 10 participants. For the first participant in the experiment, the measured timing difference was 24 milliseconds for the control trials and 224 and 324 milliseconds for the delayed trials. The difference in the delay for this participant's trials was found to be related to a difference in the delay coded in the experimental scripts.

The 95% credible intervals and mean estimates for f_o modulation extent, rate, and jitter by condition are presented in Figure 8 and Table 4. Contingent on the data and model, there was no compelling evidence that delayed auditory feedback affected the acoustical measures of vibrato.

DISCUSSION

The purpose of this study was to further investigate the effects of masked and delayed auditory feedback on the extent and rate of f_o modulation in classically-trained singers producing vibrato. This investigation was needed to clarify the role of auditory feedback in controlling vibrato due to the inconsistent findings across previous studies with small samples of participants and limited analyses. Bayesian modeling with data from ten classically-trained singers revealed that masking auditory feedback with pink noise increased the extent of f_o modulation relative to control trials with unmasked auditory feedback. This finding was

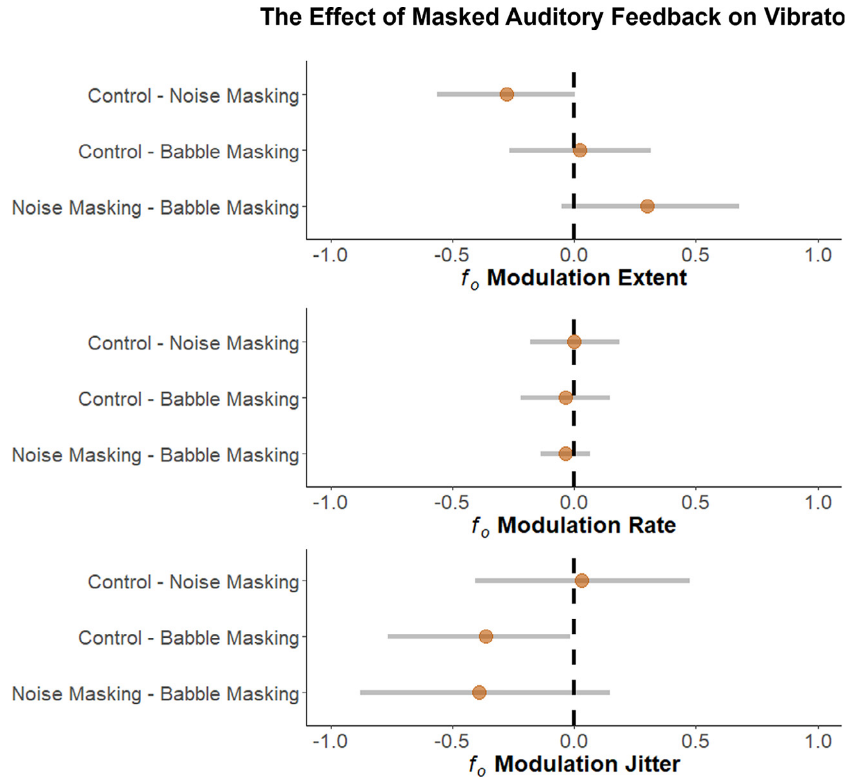

FIGURE 3. Mean estimate and 95% credible interval for f_o modulation extent (upper panel), f_o modulation rate (middle panel), and f_o modulation jitter (lower panel) for the masked auditory feedback experiments. Contrasts between each condition are listed in each panel. For the contrasts, an overlap with the zero line indicated a lack of compelling evidence for an effect.

consistent with our hypothesis and may indicate that, when singers cannot hear their intended extent of f_o modulation in their auditory feedback, they increase the extent of f_o modulation in an attempt to achieve the desired extent. This finding was inconsistent with the lack of effect of noise masking on the extent of f_o modulation in the singer studied by Ward and Burns[20] and with the predicted effect of reduced sensory feedback gain in the original reflex-resonance model[6] and the expanded reflex-resonance model.[17]

The inconsistency between the observed effect of noise masking and the predicted effect of reducing the gain of the sensory response may be related to the reflex-resonance model being based primarily on studies of somatosensory feedback, which used mechanical perturbation of the larynx with normal voice auditory feedback and computational modeling. Larson, Altman, Liu, and Hain[33] suggested that there may be linear or non-linear interactions of somatosensory and auditory feedback, wherein alteration of somatosensory feedback may oppose auditory feedback responses or adjust the gain of responses to auditory feedback. As such, it is possible that the combination of altered somatosensory feedback

TABLE 3.
Model Output for the Masked Auditory Feedback Experiments. Mean Estimate and 95% Credible Interval are Presented for the BHRM on the Effect of Masked Auditory Feedback on Each Measure of Vocal Vibrato. Rhat is Reported as an Indication of Model Convergence (at Convergence, Rhat is Around 1.00). The Posterior Probability that the Contrast Coefficients are Less than or Greater than Zero is also Presented. The Contrasts with Compelling Evidence for an Effect are in Bold.

Response	Term	Estimate	Lower	Upper	Rhat	Pr (β<0), Pr (β>0)
f_o Modulation extent	**Control - noise**	*−0.28*	*−0.56*	*0.01*	*1.00*	(0.98, *0.02*)
f_o Modulation extent	Control - babble	0.02	−0.27	0.32	1.00	(0.43, 0.57)
f_o Modulation extent	Noise - babble	0.30	−0.05	0.68	1.00	(0.05, 0.95)
f_o Modulation rate	Control - noise	0.00	−0.18	0.19	1.00	(0.50, 0.50)
f_o Modulation rate	Control - babble	−0.04	−0.22	0.15	1.00	(0.66, 0.34)
f_o Modulation rate	Noise - babble	−0.04	−0.14	0.06	1.00	(0.78, 0.22)
f_o Modulation jitter	Control - noise	0.03	−0.41	0.48	1.00	(0.43, 0.57)
f_o Modulation jitter	**Control - babble**	*−0.36*	*−0.77*	*−0.01*	*1.00*	(0.98, *0.02*)
f_o Modulation jitter	Noise - babble	−0.39	−0.89	0.15	1.00	(0.94, 0.06)

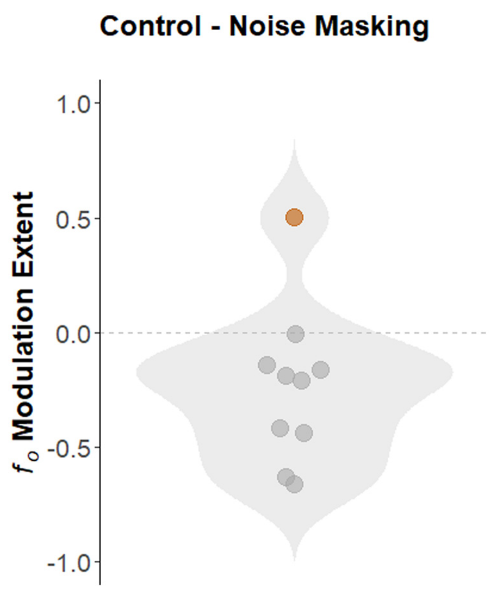

FIGURE 4. Observed differences in f_o modulation extent between control trials and noise-masked trials. Each dot represents a participant. Grey dots indicate values with higher f_o modulation extent for noise-masked trials compared with control trials. The overall negative difference reflects an increase in f_o modulation extent for noise-masked trials. The grey density shape represents the probability density along the measurement.

and typical auditory feedback in previous studies affected control of f_o in a different way than typical somatosensory feedback and altered auditory feedback in the current experiment. In addition, the findings may

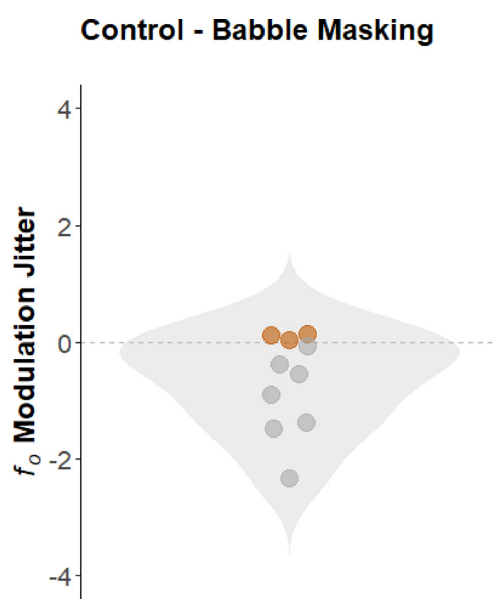

FIGURE 5. Observed differences in f_o modulation jitter between control trials and babble-masked trials. Each dot represents a participant. Grey dots indicate values with higher f_o modulation jitter for babble-masked trials compared with control trials. The overall negative difference reflects an increase in f_o modulation jitter for babble-masked trials. The grey density shape represents the probability density along the measurement.

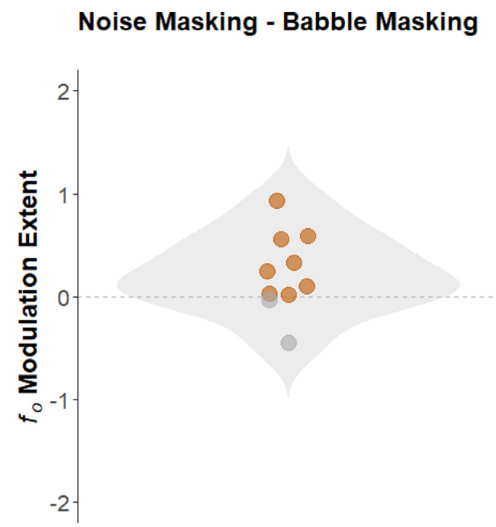

FIGURE 6. Observed differences in f_o modulation extent between noise-masked trials and babble-masked trials. Each dot represents a participant. Orange dots indicate values with higher f_o modulation extent for the noise-masked trials compared with the babble-masked trials. The overall positive difference reflects an increase in f_o modulation extent for noise-masked trials. The grey density shape represents the probability density along the measurement.

have been inconsistent with the expanded reflex-resonance model because the model was based on a typical speaker and a speaker with vocal tremor related to multiple sclerosis. Alternatively, because noise masking would have reduced the gain of auditory feedback in experimental trials, while amplifying voice auditory

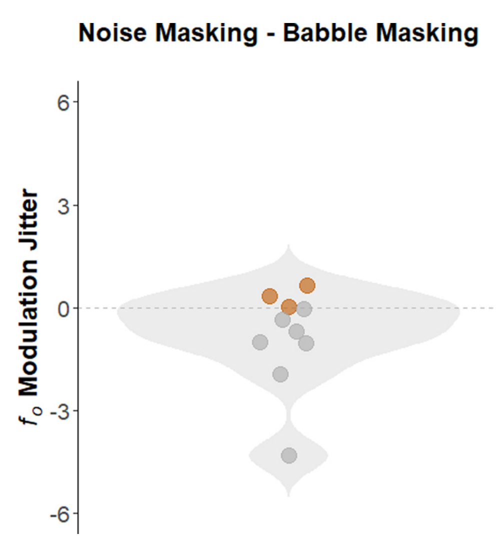

FIGURE 7. Observed differences in f_o modulation jitter between noise-masked trials and babble-masked trials. Each dot represents a participant. Grey dots indicate values with lower f_o modulation jitter for the noise-masked trials compared with the babble-masked trials. The overall negative difference reflects an increase in f_o modulation jitter for babble-masked trials. The grey density shape represents the probability density along the measurement.

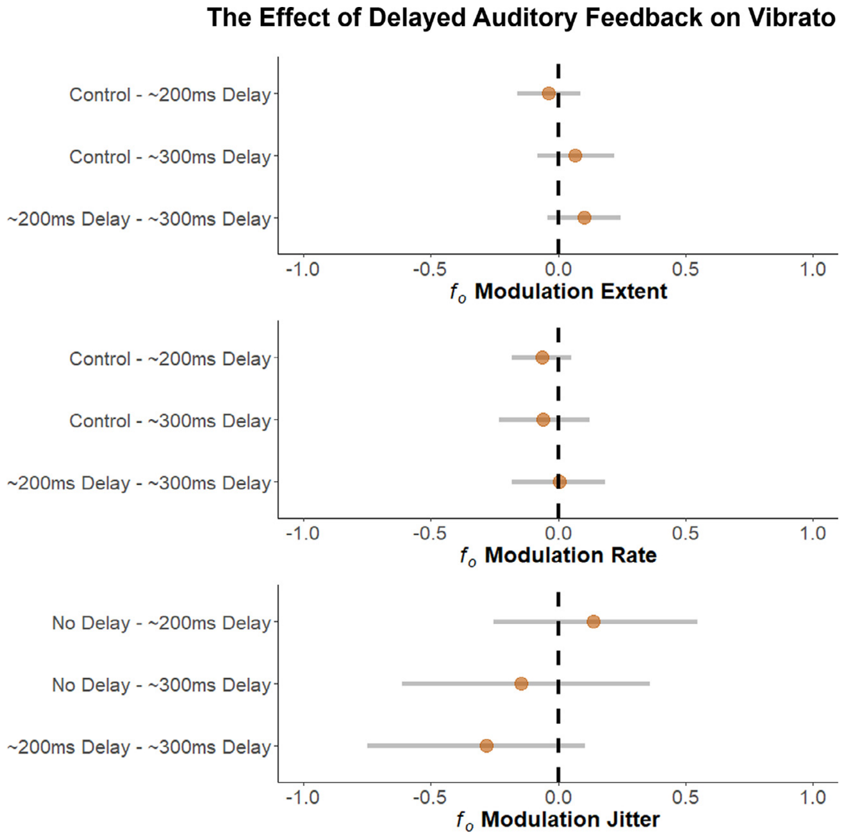

FIGURE 8. Mean estimate and 95% credible interval for f_o modulation extent (upper panel), f_o modulation rate (middle panel), and f_o modulation jitter (lower panel) for the delayed auditory feedback experiment. Contrasts are listed for each figure in each panel between each condition. For the contrasts, an overlap with the zero line indicates a lack of compelling evidence for an effect.

feedback in control trials would have increased the gain of auditory feedback, the current study may have produced a different effect than Brajot and Neiman[17] found with amplified auditory feedback only.

Although masking auditory feedback with pink noise increased the extent of f_o modulation in the current study, there was no compelling evidence that masking auditory feedback with multi-talker babble affected the extent of modulation. This contradicted our hypothesis that babble masking would have a larger effect on f_o modulation because it would not only mask the air-conducted feedback but might also distract participants from their sensory feedback. It is possible that babble did not adequately mask auditory feedback because the intensity of the babble masking varied between 76-84 dB SPL, while the intensity of the noise was consistently 80 dB SPL. It is also possible that participants habituated to the babble across trials because the same multi-talker recording was repeated for all experimental trials.

The current study revealed that there was no effect of masking auditory feedback on the rate of f_o modulation. This finding was consistent with our hypotheses, the findings of Schultz-Coloun and Battmer,[18] as cited by Shipp, Sundberg, and Doherty,[19] and predictions of the reflex-resonance models. However, babble masking did increase the jitter of f_o modulation rate, indicating that there was an increase in the variability of f_o modulation rate from one cycle of modulation to the next. Because the amplitude envelope of speech has a dominant modulation rate between 4-5 Hz (see[34] for review), and three speakers were talking simultaneously in the multi-talker babble recordings, the increase in jitter of f_o modulation may indicate that an irregular modulation of auditory feedback affected cycle-to-cycle periodicity of f_o modulation.

The current study also revealed that there was also no effect of delayed auditory feedback on extent of f_o modulation, consistent with our hypotheses, the findings of Shipp, Sundberg, and Doherty,[19] and the predictions of the reflex-resonance models. This finding was inconsistent with the findings of Deutsch and Clarkson,[21] who showed that the extent of f_o modulation increased as the delay in auditory feedback increased. The inconsistent findings may be related to differences in the imposed delays. That is, Deutsch and Clarkson[21] induced delays of 91, 197, 366, and 548 milliseconds, which differed from the induced delays in the current study and in Shipp, Sundberg, and Doherty.[19]

Finally, the current study revealed that there was no effect of delayed auditory feedback on the rate of f_o modulation. This finding was inconsistent with our hypothesis that delaying the auditory feedback would alter timing of

TABLE 4.
Model Output for the Delayed Auditory Feedback Experiments. Mean Estimate and 95% Credible Interval are Presented for the BHRM on the Effect of Delay Auditory Feedback on Each Measure of Vocal Vibrato. Rhat is Reported as an Indication of Model Convergence (at Convergence, Rhat is around 1.00). The Posterior Probability that the Contrast Coefficients are Less than or Greater than Zero is Also Presented. There were no Contrasts With Compelling Evidence for an Effect.

Response	Term	Estimate	Lower	Upper	Rhat	Pr (β<0), Pr (β>0)
f_o Modulation extent	Control – 200 ms	−0.04	−0.16	0.09	1.00	(0.75, 0.25)
f_o Modulation extent	Control – 300 ms	0.06	−0.08	0.22	1.00	(0.16, 0.84)
f_o Modulation extent	200ms – 300 ms	0.10	−0.04	0.25	1.00	(0.07, 0.93)
f_o Modulation rate	Control – 200 ms	−0.06	−0.18	0.05	1.00	(0.87, 0.13)
f_o Modulation rate	Control – 300 ms	−0.06	−0.23	0.12	1.00	(0.76, 0.24)
f_o Modulation rate	200ms – 300 ms	0.01	−0.18	0.19	1.00	(0.47, 0.53)
f_o Modulation jitter	Control – 200 ms	0.13	−0.25	0.55	1.00	(0.23, 0.77)
f_o Modulation jitter	Control – 300 ms	−0.14	−0.61	0.36	1.00	(0.73, 0.27)
f_o Modulation jitter	200ms – 300 ms	−0.28	−0.75	0.10	1.00	(0.93, 0.07)

the auditory-motor response, thereby changing the timing of the reflexive motor response and the rate of f_o modulation. The finding was also inconsistent with the findings of Deutsch and Clarkson,[21] who saw that increasing delays increased the rate of f_o modulation in 13 singers, and Shipp, Sundberg, and Doherty,[19] who saw that delays of 120, 300, and 500 milliseconds increased the rate of f_o modulation in three singers, while delays of 200 and 400 milliseconds did not affect the rate of f_o modulation. Although Shipp, Sundberg, and Doherty[19] suggested that delays of 200 and 400 milliseconds did not affect the rate of f_o modulation because they aligned with the singers' typical rates of f_o modulation, the singers reportedly had f_o modulation rates of 5.3, 5.7, and 5.8 Hz in the control trials. Therefore, delays of 189, 175, 172 milliseconds respectively (or integer multiples of these delays) would have been required to maintain an in-phase relationship with the participants' rate of f_o modulation.

The cross-correlation analyses in the current study revealed that the measured timing difference between the microphone signal and the headphone amplifier output signal was 224-240 and 324-340 milliseconds for the delayed trials. It should be noted that the measured timing difference between the recorded microphone signal and the recorded headphone amplifier output signal did not capture the additional input hardware delay, which was probably less than 10 milliseconds based on Kim, Wang, and Max.[35] Therefore, the induced delays were likely closer to 234-250 milliseconds and 334-350 milliseconds, which would correspond to f_o modulation rates of 4 Hz and 3 Hz respectively. With two participants having an f_o modulation rate close to 4 Hz, four participants having an f_o modulation rate close to 4.5 Hz, three participants having an f_o modulation rate close to 5 Hz, and one participant having an f_o modulation close to 6 Hz, delays of about 250 milliseconds, 225 milliseconds, 200 milliseconds, and 165 milliseconds would have been required to align the phase of modulation in the microphone signal and headphone signal for all participants. For future studies, the delay between the voice output and the auditory input should be measured using the procedures described by Kim, Wang, and Max,[35] and the duration of the induced delay should be aligned with each participant's typical rate of f_o modulation. Furthermore, because experimental hardware and software also induced delays in the control trials, future experiments should use normal auditory feedback for the control trials.

CONCLUSIONS

Bayesian modeling with data from ten classically-trained singers producing vibrato revealed that reducing the gain of auditory feedback with pink noise increased the extent of f_o modulation, and reducing the gain of auditory feedback with multi-talker babble increased the variability of the f_o modulation rate (ie, jitter of f_o modulation). Reducing the gain of auditory feedback did not affect the average rate of f_o modulation. Altering the gain of auditory feedback with imposed delays did not affect the average extent or rate of f_o modulation. These findings have implications for current reflex-resonance models of vocal vibrato and indicate that control of vibrato is affected by the gain of auditory feedback but may not be affected by the timing of auditory feedback.

Acknowledgments

We would like to thank Brad Story, PhD. for his previous assistance with the data analysis scripts and Melanie Looper and Elaina Derrick for their assistance with data analysis.

REFERENCES

1. Prame E. Measurements of the vibrato rate of ten singers. *J Acoust Soc Am*. 1994;96:1979–1984.
2. Shipp T, Leanderson R, Sundberg J. Some acoustic characteristics of vocal vibrato. *J Res Sing*. 1980;4:18–25.
3. 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.
4. Seashore CE. The vibrato. *University of Iowa Studies in the Psychology of Music*. 1. Iowa City, IA: University Press; 1932.
5. Nix J, Perna N, James K, et al. Vibrato rate and extent in college music majors: a multicenter study. *J Voice*. 2016;30. 756. e731-756. e741.

6. Titze IR, Story B, Smith M, et al. A reflex resonance model of vocal vibrato. *J Acoust Soc Am*. 2002;111:2272–2282.
7. 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.
8. Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*. 2006;96:280–301.
9. Lester-Smith RA, Daliri A, Enos N, et al. The relation of articulatory and vocal auditory−motor control in typical speakers. *J Speech Lang Hear Res*. 2020;63:3628–3642.
10. Burnett TA, Freedland MB, Larson CR, et al. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am*. 1998;103:3153–3161.
11. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am*. 2000;108:1246–1251.
12. Zarate JM, Zatorre RJ. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annal New York Acad Sci*. 2005;1060:404–408. https://doi.org/10.1196/annals.1360.058.
13. Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*. 2008;40:1871–1887. https://doi.org/10.1016/j.neuroimage.2008.01.026.
14. 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.
15. Leydon C, Bauer JJ, Larson CR. The role of auditory feedback in sustaining vocal vibrato. *J Acoust Soc Am*. 2003;114:1575–1581.
16. Brajot F-X, Lawrence D. Delay-induced low-frequency modulation of the voice during sustained phonation. *J Acoust Soc Am*. 2018;144:282–291.
17. Brajot F-X, Neiman AB. Vocal wow in an adapted reflex resonance model. *J Acoust Soc Am*. 2020;147:1822–1833.
18. Schultz-Coloun H, Battmer R. Quantitative evaluation of vibrato in singers. *Folia Phoniatr*. 1981;33:1–14.
19. Shipp T, Sundberg J, Doherty ET. The effect of delayed auditory feedback on vocal vibrato. *J Voice*. 1988;2:195–199.
20. Ward D, Burns E. Singing without auditory feedback. *J Res Sing*. 1978;1:4–44.
21. Deutsch JA, Clarkson JK. Nature of the vibrato and the control loop in singing. *Nature*. 1959;183:167–168.
22. 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.
23. Tumber AK, Scheerer NE, Jones JA. Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *PLoS One*. 2014;9: e109968.
24. Ramig LA, Shipp T. Comparative measures of vocal tremor and vocal vibrato. *J Voice*. 1987;1:162–167.
25. Koda J, Ludlow CL. An evaluation of laryngeal muscle activation in patients with voice tremor. *Otolaryngol Head Neck Surg*. 1992;107:684–696.
26. Lester-Smith RA, Kim JH, Hilger A, et al. Auditory-motor control of fundamental frequency in Vocal Vibrato. *J Voice*. 2021.
27. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing *[computer program]*. Vienna, Austria: R Foundation for Statistical Computing; 2021. Version 4.0.5.
28. RStudio. RStudio: Integrated Development Environment for R *[computer program]*. Boston, MA: RStudio; 2021. PBC.
29. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. *J Statistical Softw*. 2017;76:1–32.
30. Bürkner P. Advanced Bayesian multilevel modeling with the R package brms. *R J*. 2018;10:395–411. https://doi.org/10.32614/RJ-2018-017.
31. Nalborczyk L, Batailler C, Lúvenbruck H, et al. An introduction to Bayesian multilevel models using brms: a case study of gender effects on vowel variability in standard Indonesian. *J Speech Lang Hear Res*. 2019;62:1225–1242.
32. Barr DJ, Levy R, Scheepers C, et al. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Memory Lang*. 2013;68:255–278.
33. Larson CR, Altman KW, Liu H, et al. Interactions between auditory and somatosensory feedback for voice F 0 control. *Experiment Brain Res*. 2008;187:613–621.
34. Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. *Nat Rev Neurosci*. 2020;21:322–334.
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 Hearing Res*. 2020;63:2522–2534.