Modulation of auditory-vocal feedback control due to planned changes in voice f_0

Jason H. Kim and Charles R. Larson^{a)}

Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA

(Received 4 June 2018; revised 11 February 2019; accepted 26 February 2019; published online 25 March 2019)

Previous studies have demonstrated that voice fundamental frequency (f_0), or pitch, relies on auditory feedback to monitor and correct for errors in production. When voice-pitch auditory feedback is unexpectedly perturbed, individuals typically produce a compensatory change in f_0 that opposes the direction of the pitch-perturbation. Studies comparing steady vowel vocalizations and speech tasks have demonstrated task-dependent modulation of the compensatory response, but the effects of planning to volitionally change f_0 during active vocalization have yet to be explored. Ten musicians and ten non-musicians were asked to perform two vocal tasks. Both tasks started off at a conversational f_0 . In one task, pitch-shifted feedback was presented when the participants were planning to hold f_0 constant (steady f_0), and in the other, feedback was shifted while participants were in the planning stage prior to raising f_0 (raised f_0) from a steady state. Acoustical analyses of f_0 were performed to measure the peak magnitude and latency of both the compensatory response as well as the voluntary f_0 change. Results showed that planning to change pitch modulates the mechanisms controlling feedback-based error correction of f_0 , and musicality affects how individuals incorporate modulations in auditory feedback with the feedforward plans to increase voice f_0 .

© 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5094414>

[ZZ]

Pages: 1482–1492

I. INTRODUCTION

Vocalization is a complex activity that is important for speech, singing, and emotional expression. An important aspect of vocalization is the ability to control one's fundamental frequency (f_0) (perceived as pitch) or loudness so that musical expression, linguistic intent, or emotional context can be properly communicated. Motor control of voice f_0 and loudness involves the coordination of respiratory and laryngeal muscles to drive the vocal folds into vibration and produce the desired pitch. Once vocalization begins, individuals monitor auditory feedback for errors in their voice f_0 or loudness and automatically produce compensatory responses if errors are detected (Burnett *et al.*, 1998; Scheerer and Jones, 2017; Larson and Robin, 2016; Bauer *et al.*, 2006; Hain *et al.*, 2000; Behroozmand *et al.*, 2012). This property of the audio-vocal system has been likened to a negative feedback control system (Hain *et al.*, 2000). Although most vocal responses are compensatory in nature, responses that follow the direction of the pitch or loudness change are also reported (Burnett *et al.*, 1998; Franken *et al.*, 2018; Hain *et al.*, 2000; Behroozmand *et al.*, 2012). The latter types of responses may be considered to operate as a feedforward system and would seem to represent the intention of the speaker to achieve a specific form of vocalization (Patel *et al.*, 2014). Both forms of feedback are necessary for optimal vocal control: the desire to achieve a particular type of vocalization (pitch, loudness, duration, etc.) and the correction for errors in production.

Much of our understanding of vocal control is based on studies that perturbed pitch auditory feedback during steady vowel vocalizations (Burnett *et al.*, 1998; Larson *et al.*, 2000; Liu and Larson, 2007; Liu *et al.*, 2011; Korzyukov *et al.*, 2012). However, these tasks are ultimately not representative of the dynamic control necessary for monitoring the variable pitch fluctuations that are natural to speech and singing.

Some previous studies have attempted to address these shortcomings by investigating how compensatory responses are modulated by different tasks (Natke and Kalveram, 2001; Burnett and Larson, 2002; Natke *et al.*, 2003; Chen *et al.*, 2007; Liu *et al.*, 2009). These studies suggest that different vocal plans involved in feedforward control can lead to alternative sensory expectancies from those of steady vocalizations. Singing tasks involving nonsense syllables have been shown to elicit larger compensatory response magnitudes than speech tasks, presumably due to the greater precision of pitch control needed in the former (Natke *et al.*, 2003). Chen *et al.* (2007) further explored this issue of task-dependency by comparing compensatory f_0 responses during steady vocalizations to those in a speech task ("You know Nina?"). They found larger compensatory response magnitudes in the speech task, which involved an increase in f_0 towards the end of the phrase in order to impart an inquisitive linguistic intent. Interestingly, the authors noted an interaction between vocal task and stimulus direction, which suggested that larger compensatory responses were produced when the perceived feedback error (downward or decreased pitch) conflicted with the direction of the intended feedforward speech intonation plan (raised pitch). As a whole, the

^{a)}Electronic mail: clarson@northwestern.edu

literature suggests that the neural system constantly monitors auditory feedback of vocal motor production to facilitate feedforward control of suprasegmental features in planned utterances.

In contrast, other studies have investigated how volitional intent interacts with anticipated modifications to pitch auditory feedback (Hain *et al.*, 2000; Patel *et al.*, 2014). When individuals are instructed to raise or lower their f_0 in response to changes in auditory feedback, they tend to involuntarily produce a compensatory f_0 response prior to their voluntary f_0 change (Hain *et al.*, 2000). On the other hand, when subjects are instructed to oppose the direction of a change in voice pitch feedback, they have a tendency to produce small following responses, which then delay the onset of the opposing response (Patel *et al.*, 2014). Together, these two studies demonstrate an interaction between volitional vocal control and anticipated changes in feedback. However, it is important to note that these previous works required participants to actively change voice f_0 while responding to large magnitude, extended pitch-shifts ($\geq \pm 200$ cents, > 500 ms) in their auditory feedback, fully aware that feedback was intentionally manipulated by the experimenters. In comparison to most pitch-shift studies, which used small magnitude and short-duration ($\leq \pm 100$ cents, 200 ms) pitch-shifts, large and extended-duration pitch-shifts are more salient changes in auditory feedback for participants.

Similarly, some previous studies have explored how vocal control is influenced by both volitional intent and the salience of pitch-shifts by examining musically trained and non-musically trained populations. In regards to pitch-shift salience, trained singers demonstrated the ability to ignore extended-duration 200 cent shifts, but they were unable to inhibit responses to extended-duration 25 cent shifts (Zarate *et al.*, 2010). However, when asked to ignore extended 200 cent shifts in a similar experiment, non-musicians were unable to suppress the responses, in comparison to a group of trained singers (Zarate and Zatorre, 2008). In contrast, other studies have demonstrated that musically trained individuals, compared to non-musicians, have larger compensatory and neural responses to perturbations in voice pitch auditory feedback during steady voice f_0 production, suggesting a greater ability to detect and correct for perceived errors (Behroozmand *et al.*, 2014; Parkinson *et al.*, 2014; Sturgeon *et al.*, 2015). The significant difference between these seemingly contradictory results is that these latter studies used small magnitude and short-duration pitch-shifts. Together, these results suggest that the influence of volitional intent on vocal control appears to be associated with the salience of a pitch-shift in a vocal task and have a significant interaction with the musical training of an individual. Moreover, musicality and musical training have been correlated with various abilities such as greater pitch discrimination (Nardo and Reiterer, 2009), and auditory-motor changes can result from long-term practice and expertise demonstrated by musicians (Kleber *et al.*, 2010; Halwani *et al.*, 2011). Due to the overlap of neural substrates involved in singing and vocalization (Zarate, 2013), it stands to reason that musical training would also impact how one plans the movements required for voice f_0 control and adjusts the

planned movements accordingly to unanticipated mismatches between feedback and intent.

To further investigate the role of planning in feedforward and feedback control of vocal pitch, the present study implemented a behavioral paradigm which asked participants to vocalize while listening to their own voice auditory feedback and planning to raise their voice f_0 on cue. The feedback contained pitch-shifts that occurred immediately prior to implementing the *planned* changes in f_0 . The goal of this experiment was to explore whether feedback-modulated voice f_0 control for unanticipated errors in feedback differed prior to initiating voluntary changes in f_0 , specifically amongst feedforward plans to hold voice f_0 steady (Steady f_0) or increase f_0 (Raised f_0). This study would add to the literature by investigating ongoing auditory-vocal feedback control related to less-salient errors (small magnitude, short-duration) prior to engaging in planned, volitional vocal control, devoid of potential articulatory or suprasegmental influences involved in speech production. As proposed by Chen *et al.* (2007), audio-vocal control is sensitive to the direction of perceived errors when they conflict with intended intonation patterns. In the current study, it was proposed that if similar differences could be found between compensatory responses during steady vocalizations and those prior to volitional f_0 increases, greater support would be provided for a general auditory-vocal sensitivity to planned changes in feedforward voice control. However, if there are no differences between tasks or a different pattern of compensatory responses emerges, results would suggest that previous reports were specific for prosodic or linguistic contexts. The present study also explored the role of musicality in planned changes of auditory-vocal control by comparing responses from musically trained and non-trained individuals. Musicians were predicted to produce compensatory responses of larger magnitudes and shorter latencies prior to voluntary f_0 changes compared with non-musicians. These results would indicate that greater musical experience can provide heightened sensitivity to the errors in feedback prior to planned f_0 changes, which would allow for more successful productions of the intended change in f_0 . Larger responses for musicians compared to non-musicians in the steady f_0 task would indicate that musicians are more sensitive to perturbations when their intent is to maintain a steady f_0 .

II. METHODS

A. Participants

Twenty American-English speaking adults (12 female and 8 male; age range: 18–32 years) volunteered as participants. All participants passed a bilateral pure-tone hearing screening [octave intervals from 250 to 8000 Hz at 20 dB hearing level (HL)] and denied a history of voice, speech, hearing, or neurological disorders. Half of the participants ($N = 10$, seven female) were required to have minimal to no musical training (“non-musicians,” or NM), which was defined as four years or less of musical instruction (estimated average = 1 year; range: 0–4 years). Additionally, NM participants did not have any musical instruction or practice at

least three years before the start of the present study. The other participants ($N=10$, five female) were required to have significant musical training, which was defined as having five or more years of musical instruction or continued practice (minimum of $1 \times$ a week) (estimated average = 9 years; range: 5–20 years). Instrumental training among participants ranged from piano, guitar, violin, clarinet, and saxophone. None of the participants reported familiarity with or instruction in tonal languages. Twenty subjects were recruited because previous pitch-shift studies using mixed designs (between-groups and within-subjects comparisons) have employed similar numbers of participants and have observed robust effects (e.g., [Korzyukov et al., 2015](#)). All participants provided informed consent approved by the Northwestern University Institutional Review Board.

B. Instrumentation

Participants were seated in a sound-attenuated booth (IAC model 1201) and wore Sennheiser headphones (model HMD 280) with an attached microphone placed 1 in. from the corner of the mouth. Participants monitored their vocal intensity using a Dorrrough Loudness monitor in order to maintain a comfortable intensity of 70 dB sound pressure level (SPL). The microphone signal was digitized with a MOTU Ultralite Mk3 audio interface, and audio output was fed back to the headphones, amplified with an Aphex Headpod 4 to provide a cumulative 10 dB gain from the voice output to minimize the influence of air- and bone-conducted voice feedback. A Brüel and Kjær sound level meter (type 2250), binaural in-ear microphones (type 4101-A), and a prepolarized free-field microphone (type 4189) were used to calibrate the gain between voice output and feedback channels with a 1 kHz sinusoidal pure tone.

A custom-designed, non-commercial audio-visual interface (PitchPresent4) developed in our lab was used to record and digitize voice output, voice auditory feedback, and transistor-transistor-logic (TTL) pulses. PitchPresent4, is a software interface written in c++ and Qt. Source code from the smbPitchShift system ([Bernsee, 1999](#)), was used to provide pitch shifts of online voice auditory feedback and control parameters for f_0 perturbations, including perturbation direction, magnitude, and timing. The PitchPresent4 interface also displayed visual prompts to participants and simultaneously produced transistor-transistor logic (TTL) pulses to signify the onset of the pitch shift stimulus. The PitchPresent4 system had an estimated latency of 20 ms between production and playback of the voice feedback.

C. Procedures and experimental design

At the beginning of each trial, the visual monitor displayed “Say ‘ahh’” to notify participants when to prepare for and begin vocalizing. For both the steady and raised f_0 tasks, participants were instructed to vocalize /a/ at a comfortable, but stable, conversational pitch and loudness. It was heavily emphasized to participants that they start with the same f_0 pitch, regardless of the task. When triggered by voice onset, the monitor displayed a sequence of visual symbols arranged as a countdown leading to a possible increase in voice pitch

(“3-2-1-GO”) (Fig. 1). Each number in the countdown sequence was presented for 500 ms. This visual prompt was used in order to help make the cadence for the vocal task uniform from trial to trial and across subjects, as well as to prevent the stimulus from being presented during the rise in pitch. The visual representation consisted of five horizontal lines, with a thicker medial line indicating their conversational f_0 level. In the steady f_0 task, the visual representation included a circle on the medial line, which instructed participants to ignore the countdown and keep their f_0 at their conversational pitch for the entire trial. In the raised f_0 task, the visual representation had the circle situated on the line above their conversational pitch line, which instructed participants to produce a steady voice f_0 followed by an upward change in pitch when “GO” appeared.

A 200 ms long pitch perturbation (± 50 cents) occurred 300 ms before “GO” appeared on the visual display, allowing enough time for a compensatory response to occur before participants volitionally raised f_0 ([Hain et al., 2000](#); [Patel et al., 2014](#)). Control trials did not have pitch perturbations. Participants were instructed to raise their f_0 as quickly as possible when “GO” appeared (2.5 s after voice onset), but not before the cue. Once they had raised their f_0 , participants were asked to maintain the raised f_0 level until the end of the trial (1.5 s after “GO” appeared).

The steady and raised f_0 tasks were randomly distributed across 180 trials and presented in 6 blocks of 30 trials with 1–5 min rest breaks between each block. Each trial was approximately 6 s long, with 4 s of vocalization and 2 s of rest. To successfully complete the tasks, participants had to pay close attention to both the visual representations of the vocal tasks as well as the countdown cues. Participants were informed before the start of the experiment that they could rest between trials for as long as they needed, as long as they did not trigger the microphone.

D. Data analysis

1. Voice response pre-processing

A custom-designed non-commercial software (PitchBrowse4), developed in our lab, was used to interface with the audio file output of PitchPresent4. PitchBrowse4 is a software interface written in c++ and employs an autocorrelation method in Praat ([Boersma and Weenink, 2017](#)) to convert voice audio signals into voice f_0 contours. PitchBrowse4 was used to segment the f_0 contours into epochs of 1200 ms duration (200 ms pre-pitch shift and 1000 ms post-pitch shift). Each pitch-shifted segment or control segment were then sorted by direction of the perturbation and vocal task. The signals were converted from Hz to cents using the formula, cents = $1200 \times \log_2(f_2/f_1)$, where f_2 denotes the f_0 values every 5 ms for a single post-shift epoch and f_1 is the mean f_0 magnitude of the corresponding pre-shift epoch. After the segmentation and f_0 -to-cent conversion, PitchBrowse4 outputted a text file detailing the f_0 cent contour data and their associated condition labels (e.g., +50 cents steady f_0 task).

The text file outputs from PitchBrowse4 were then uploaded to an in-house, non-commercial online analysis tool, Speech Analysis Web Analysis. Speech Analysis is a

FIG. 1. Visual prompts for the raised (top) and steady f_0 (bottom) vocal tasks. The prompts were displayed on a computer monitor in the sound booth. When “Say ‘ahh’” appeared, participants were instructed to start vocalizing with a steady, conversational pitch for both tasks. The displays switched to the task specific displays after voice onset. In the raised f_0 task, participants were instructed to increase their f_0 as soon as “GO” appeared and maintain that level until “Stop” appeared. In the steady f_0 task, participants maintained their conversational pitch with a steady vocalization for the entirety of the trial. A pitch perturbation was applied 300 ms before “Go” appeared for a duration of 200 ms. Participants continued vocalizing for 1.5 s after “Go” appeared.

web interface written in JavaScript and HTML that housed the data in a MySQL database with php scripts. Speech Analysis allows f_0 cent contours to be viewed as individual trials, waterfall plots of trials sorted into conditions (vocal task, steady and raised f_0 ; perturbation direction, +50 vs –50 vs no perturbation), and grand averages of sorted trials. Within Speech Analysis, individual segments with perturbations were then further sorted as a compensatory or following response type, on a trial by trial basis, in relation to the direction of their response to the stimulus. As performed in previous studies (Behroozmand *et al.*, 2012), the mean f_0 value from 100 to 300 ms (post-stimulus onset) was compared to the mean f_0 value from –200 to 0 ms (pre-stimulus onset) for each individual trial that contained a pitch perturbation. In contrast to the study by Behroozmand *et al.* (2012), the present study used time windows of 200 ms for measuring both the pre- and post-stimulus onset means in order to compare f_0 averages across equal lengths of time. Although Behroozmand *et al.* (2014) chose a post-stimulus onset window of 50 to 250 ms, the present study used the time window from 100 to 300 ms in order to more accurately contain the average peak response latency of the compensatory response, which consequently led to less ambiguous identification between opposing and following responses. Table I demonstrates that in the raised f_0 trials, peak compensatory response latencies occurred between 250 and 300 ms post-stimulus onset. Although peak compensatory response latencies occurred between 280 and 330 ms in the steady f_0 trials, the visual instruction to volitionally change f_0 was given at 300 ms in the raised f_0 trials. Therefore, 300 ms was selected as the upper limit for the time window in order to prevent the possibility of volitional f_0 changes from influencing the

response direction identification. Compared to the pre-stimulus onset mean, a trial with a larger post-stimulus onset mean would be categorized as an “upward” response, and a trial with a smaller post-stimulus onset mean would be categorized as a “downward” response. Responses that went in the opposing direction of the pitch perturbation were considered “compensatory” responses [e.g., a downward response to an upward (+50 cent) pitch-shift], whereas those that mimicked the direction of the perturbation were considered “following” responses [e.g., an upward response to an upward (+50 cent) pitch-shift].

Visual inspection of trials was performed for each subject in order to remove trials that met the following exclusion criteria: trials containing flat, horizontal f_0 contours from inaccurate pitch tracking or interrupted vocalizations; trials containing noise from coughs or late starts; or artifacts from inaccurate pitch tracking. Trials were excluded if participants changed their f_0 before the “GO” signal or if participants did not raise their f_0 during a raised f_0 task. In the raised f_0 task, a proportion of trials were excluded primarily because participants produced early voluntary changes that occurred before the “GO” signal. It is possible that these

TABLE I. Average peak latency (ms) of compensatory responses to pitch perturbations (± 50 cents) in the steady and raised f_0 tasks. Standard error of the mean is indicated in parentheses.

	Steady f_0		Raised f_0	
	+50c	–50c	+50c	–50c
Overall	317.5 (9.87)	289.25 (10.53)	266.5 (9.47)	280 (9.37)
Non-musicians	323 (16.35)	294.5 (17.90)	259 (13.10)	269.5 (13.34)
Musicians	312 (11.72)	284 (11.90)	274 (13.96)	290.5 (12.96)

“early volitional” changes may have reflected a combination of reflexive responses and voluntary changes to f_0 as observed by Hain *et al.* (2000). However, it was not possible to definitively label them as combined responses in the present study. Similarly, Patel *et al.* (2014) also demonstrated that short-latency following responses can be produced prior to volitional changes, but due to the limitations of using only a raised f_0 vocal task, it was not always possible to identify the direction of the reflexive response that contributed to this early response or describe potential interactions that may have been present among the type of reflexive response, the direction of the pitch-shift, and the direction of the intended volitional change. Coupled with the fact that following responses were sometimes absent or scarce for some participants in the raised f_0 task, all trials with early volitional changes and following responses were excluded from the analysis. Likewise, if a participant incorrectly performed the raised f_0 task during a steady f_0 task, the trial was excluded. Approximately 19.72% of trials were excluded from Speech Analysis due to meeting one of the aforementioned exclusion criteria.

2. Fundamental frequency (f_0) analysis

Grand averaged waveforms for each combination of vocal task and perturbation direction were exported from Speech Analysis to IGOR PRO (v. 6.0, Wavemetrics, Inc.) as text files for response measurement. The dependent variables (peak magnitude and latency) of the *compensatory f_0 response* for each perturbation direction (+50 and −50 cents) in both vocal tasks (Steady and Raised f_0) were measured. Likewise, the magnitude and latency of the *peak voluntary f_0 change in the raised f_0 task* was measured, across perturbation conditions (+50 cents and, −50 cents, and no perturbation).

A mixed factorial repeated-measures-analysis of variance (RM ANOVA) (SPSS, IBM) was the main statistical procedure used for the behavioral analysis. The between-participants factor of group assignment was based on musicality (musicians and non-musicians). The within-subjects factors were a 2×2 design of task (steady and raised f_0) and perturbation direction (+50 cents and −50 cents). The dependent variables of the compensatory f_0 response, peak response magnitude and latency were independently assessed in separate ANOVAs. For the voluntary f_0 changes in the raised f_0 task, a 2×3 design of the same between-subjects group factor (musicality) and perturbation direction (+50 cents, −50 cents, and no perturbation) was used. All data were assessed for normality and equality of error variances using the Shapiro-Wilk test and the Levene’s test, respectively. All *post hoc* tests for significant interactions used a Bonferroni adjustment for multiple comparisons.

III. RESULTS

A. Compensatory f_0 responses to pitch-perturbed feedback prior to voluntary f_0 changes

As observed in previous studies (Burnett *et al.*, 1998; Bauer and Larson, 2003), participants produced

compensatory responses in the majority of trials with perturbations, but they did not produce responses in control trials without perturbations (Fig. 2).

Importantly, participants produced compensatory responses prior to volitionally raising their f_0 . Figures 3 and 4 show the compensatory f_0 responses in the steady f_0 and raised f_0 tasks, respectively. A mixed factorial RM ANOVA identified a main effect of vocal task on the peak response latency [$F(1,18) = 10.226$, $p < 0.01$, $\text{partial } \eta^2 = 0.362$], with responses in the raised f_0 task producing shorter peak latencies than those in the steady f_0 task (Fig. 5, Table I). An interaction of vocal task and perturbation direction further indicated that peak response latencies to +50 cent perturbations were significantly shorter in the raised f_0 task compared to the steady f_0 task ($p < 0.01$) (Table II). Response magnitudes did not differ significantly between vocal tasks. Although there were no group differences for peak compensatory response latency, musicians produced larger peak magnitudes across all trials [$F(1,14) = 9.609$, $p < 0.01$, $\text{partial } \eta^2 = 0.348$] (Fig. 6, Table II). A main effect of perturbation direction also indicated that compensatory responses to −50 cent perturbations produced larger magnitudes than those to +50 cent perturbations, regardless of the vocal task [$F(1,18) = 61.976$, $p < 0.001$, $\text{partial } \eta^2 = 0.775$] (Table II).

B. Effects of pitch perturbed feedback on voluntary f_0 changes

In the raised f_0 task, participants were asked to volitionally raise their f_0 from a steady state when visually prompted. Figure 7 compares the voluntary f_0 changes across the perturbation conditions (+50 cents, −50 cents or no perturbation) between non-musician and musician groups. Interestingly, an interaction of group and perturbation condition in voluntary f_0 change latency indicated that although there were no group differences in trials with perturbations, in control trials, musicians had shorter voluntary response latencies than non-musicians [$F(2,36) = 6.575$, $p < 0.01$, $\text{partial } \eta^2 = 0.268$] (Fig. 8, Table III). That is, in trials without perturbations, musicians reached their peak voluntary f_0 change more quickly compared to non-musicians. There were no significant findings for differences in the peak voluntary f_0 change magnitude between groups or for perturbation condition (Table IV).

IV. DISCUSSION

The present study investigated the role of planning in modulating feedforward and feedback control of vocal pitch. To observe the effects on feedback control, compensatory f_0 responses to pitch perturbations were compared between the steady and raised f_0 vocal tasks. Importantly, differences in response latency indicate that peak compensatory responses in the raised f_0 task occurred earlier than in the steady f_0 task. This effect was driven by the shorter compensatory response latencies to upward (+50 cent) perturbations in the raised f_0 task compared to the steady f_0 task. Overall, compensatory response magnitudes to downward (−50 cent) perturbations were larger in both tasks than to upward perturbations. When compared across musicality, musicians

FIG. 2. Comparisons with the grand averages of the steady and raised f_0 tasks in control trials (no perturbation). The dashed line indicates the time at which the "GO" visual cue was presented. (A) The steady and raised f_0 cent wave diverge at approximately 300 ms, after the visual cue to increase pitch for the raised f_0 task is presented. (B) When examining the full trace of the raised f_0 response, participants on average reached their peak voluntary f_0 change at 640.75 ms in trials without perturbations. Participants maintained f_0 levels with little variation for both the (C) steady f_0 task and (D) raised f_0 task during control trials, whereas pitch-shifts in either direction produced compensatory responses for both tasks. Error bars indicate the standard error of the mean.

produced larger compensatory response magnitudes than non-musicians overall, but there were no task differences between groups. Feedforward control was assessed in the raised f_0 task by comparing the peak voluntary f_0 change (increase) in the control condition trials (no perturbation) to trials with perturbations (+50 or -50 cent shift) prior to the volitional f_0 increase. Although there were no differences in voluntary f_0 change between perturbation directions, the results indicate that musicians reached their peak voluntary

f_0 change magnitude with shorter latencies than non-musicians in the control task while displaying no group differences in perturbed trials.

A. Modulating feedback control of voice f_0 with volitional intent

Previous research has investigated auditory-vocal feedback control by perturbing pitch during various tasks,

FIG. 3. Compensatory responses to ± 50 pitch perturbations for non-musicians (black) and musicians (grey) in the steady f_0 task. Responses to (A) +50c (upward) perturbations produced downward compensations, and responses to (B) -50c (downward) perturbations produced upward compensations. The dashed line indicates the time at which the "GO" visual cue was presented. Error bars indicate the standard error of the mean.

FIG. 4. Compensatory responses to ± 50 pitch perturbations for non-musicians (black) and musicians (grey) in the raised f_0 task. Responses to (A) +50c (upward) perturbations produced downward compensations, and responses to (B) -50c (downward) perturbations produced upward compensations. The dashed line indicates the time at which the “GO” visual cue was presented. Error bars indicate the standard error of the mean.

including singing (Burnett and Larson, 2002; Natke *et al.*, 2003), nonsense German words (Natke and Kalveram, 2001; Donath *et al.*, 2002) prolonged Mandarin tones (Jones and Munhall, 2002), Mandarin speech (Xu *et al.*, 2004; Liu *et al.*, 2009), and English speech (Chen *et al.*, 2007). Taken together, the results overwhelmingly demonstrated that auditory feedback control of voice f_0 is active in ongoing fluctuations in f_0 during a variety of voluntary voice control conditions.

Furthermore, task-dependent effects on feedback control have been investigated by comparing compensatory responses from dynamic vocal tasks to the more commonly used steady vowel vocalizations. Although task-dependent modulations were often found across studies, it has been unclear how the results synthesize as a whole in describing how compensatory feedback control of dynamic vocal or speech tasks differ from that of steady state vocal tasks. For example, when singers heard pitch perturbed feedback while producing upward pitch glissandos, their compensatory responses had smaller magnitudes and longer latencies, compared to responses in steady vowel tasks (Burnett and

Larson, 2002). On the other hand, when individuals received pitch perturbed feedback during a Mandarin speech task with a rising intonation, their compensatory responses magnitudes and latencies did not differ compared to responses from a speech task with a flat, steady intonation (Liu *et al.*, 2009). These contrasting results are possibly due to methodological differences between studies, such as the linguistic background of participants, stimulus magnitude, or the presence of speech movements. In particular, the studies have differed in how response magnitude is measured in the context of a rising slope in the f_0 trace. The compensatory responses during voice glissandos measured f_0 magnitude in relation to an extrapolated f_0 baseline (Burnett and Larson, 2002), while responses in the speech task measured the greatest f_0 magnitude from a difference wave between the perturbed and control trials (Liu *et al.*, 2009).

However, an alternative explanation may lie in the timing of the pitch perturbation. In the glissando task (Burnett and Larson, 2002), feedback was perturbed while participants were in the midst of volitionally increasing f_0 , whereas in the speech task (Liu *et al.*, 2009), feedback was perturbed during the relatively stable f_0 prior to the voluntary f_0 increase. Similar to Liu *et al.* (2009), the current study did not observe response magnitude differences between the steady and raised f_0 tasks when the perturbation was applied at a similar time prior to the voluntary rise in f_0 . Liu *et al.* (2009) did, however, include perturbations at various time points prior to the rising intonation and found that perturbations earlier in the utterance produced larger response magnitudes when compared to the perturbation proximal to the

FIG. 5. Boxplots showing compensatory response latencies (ms) for both perturbation directions (± 50 cents) in the raised (grey) and steady (white) f_0 tasks. Boxplot definitions: the middle horizontal line represents the median. The upper and lower limits of the box indicate the 75th and 25th percentiles, respectively. The length of the central rectangle represents the interquartile range (IQR). The whiskers extend to the upper and lower limits of all data points, excluding outliers. Unfilled circles indicate mild outliers ($IQR \times 1.5$). **Indicates a significant effect at the level of $p < 0.01$, adjusted for multiple comparisons.

TABLE II. Average magnitude (cents) of compensatory responses to pitch perturbations (± 50 cents) in the steady and raised f_0 tasks. Response magnitudes are expressed in absolute value. Standard error of the mean is indicated in parentheses.

	Steady f_0		Raised f_0	
	+50c	-50c	+50c	-50c
Overall	21.88 (1.76)	26.43 (2.49)	18.60 (2.08)	26.99 (1.83)
Non-musicians	17.32 (1.79)	20.19 (2.49)	13.37 (1.25)	23.04 (2.52)
Musicians	26.45 (2.28)	32.66 (3.36)	23.83 (3.26)	30.94 (2.08)

FIG. 6. Boxplots showing compensatory response magnitudes (cents) for both perturbation directions (± 50 cents) in the raised (grey) and steady (white) f_0 tasks. The response magnitudes of the musician group are displayed on the left, and the response magnitudes of the non-musician group are displayed on the right. Boxplot definitions: the middle horizontal line represents the median. The upper and lower limits of the box indicate the 75th and 25th percentiles, respectively. The length of the central rectangle represents the IQR. The whiskers extend to the upper and lower limits of all data points. **Indicates a significant effect at the level of $p < 0.01$, adjusted for multiple comparisons.

f_0 increase. Based on these results, Liu *et al.* (2009) suggested that the planning process for voluntary changes in f_0 have a critical time period earlier in speech production in which compensatory response mechanisms are more sensitive to auditory feedback.

Thus, the present study and past literature point to a potential trend for vocal utterances with a planned change in f_0 . Compensatory responses to feedback perturbations appear to be largest when perturbations occur near voice onset and gradually attenuate as perturbations are applied closer to the

FIG. 7. Voluntary f_0 increases for non-musicians (black) and musicians (grey) in the raised f_0 task. The magnitude and latency of the peak voluntary f_0 change was measured in trials with (A) +50c (upward) perturbations, (B) -50c downward perturbations, and (C) no perturbations (control). The dashed line indicates the time at which the “GO” visual cue was presented. Error bars indicate the standard error of the mean.

FIG. 8. Boxplots showing voluntary f_0 change latencies (ms) for both perturbation directions (± 50 cents) in the raised and steady f_0 tasks. The response magnitudes of the musician group are displayed in grey, and the response magnitudes of the non-musician group are displayed in white. Boxplot definitions: The middle horizontal line represents the median. The upper and lower limits of the box indicate the 75th and 25th percentiles, respectively. The length of the central rectangle represents the IQR. The whiskers extend to the upper and lower limits of all data points, excluding outliers. Unfilled circles indicate mild outliers ($IQR \times 1.5$), and filled circles indicate extreme outliers ($IQR \times 3$). **Indicates a significant effect at the level of $p < 0.01$, adjusted for multiple comparisons.

volitional f_0 change, with the smallest compensatory responses when a volitional change occurs immediately in advance or concurrently with the perturbation. This trend suggests that compensatory responses are more sensitive to perceived errors in auditory feedback near vocal onset in a steady-state utterance but gradually rely less on compensatory feedback control and more on dynamic feedforward control in advance of the planned change in f_0 . These critical periods may be particularly sensitive near vocal onset because the state of the vocal mechanism at the beginning of production would be important for determining how the rest of the vocal intent will be implemented. Future studies will be needed to address this potential holistic framework in understanding how planning integrates with feedback and feedforward vocal control.

B. Pitch perturbation direction in relation to planned volitional f_0 direction

In a previous study, [Chen et al. \(2007\)](#) perturbed pitch feedback while participants produced a phrase (i.e., “You know Nina?”) and a steady vowel to compare compensatory responses between the two tasks. [Chen et al. \(2007\)](#) found that compensations to downward perturbations produced the largest response magnitudes and shortest peak latencies, exclusively for the speech task. From these results, it was concluded that auditory-vocal control is sensitive to the

TABLE III. Average latency (ms) of peak voluntary f_0 change in the raised f_0 task for control (no perturbation) and perturbed trials (± 50 cents). Standard error of the mean is indicated in parentheses.

	Control (no perturbation)	+50 cent perturbations	-50 cent perturbations
Overall	640.75 (18.08)	634.75 (14.26)	637 (13.90)
Non-musicians	677.5 (28.99)	647 (18.93)	643.5 (22.26)
Musicians	604 (15.45)	622.5 (21.62)	630.5 (17.65)

TABLE IV. Average magnitude (cents) of peak voluntary f_0 change in the raised f_0 task for control (no perturbation) and perturbed trials (± 50 cents). Standard error of the mean is indicated in parentheses.

	Control (no perturbation)	+50 cent perturbations	-50 cent perturbations
Overall	390.31 (30.00)	385.19 (35.80)	394.84 (31.36)
Non-musicians	414.16 (49.09)	415.43 (60.09)	420.55 (53.34)
Musicians	366.46 (35.56)	354.96 (39.99)	369.14 (34.04)

direction of perceived errors when they conflict with planned intonation patterns, such as with an increase in f_0 to signal a question. To test this idea, the current study similarly predicted that compensatory responses to downward perturbations would result in larger peak magnitudes and shorter latencies, which would indicate that feedback control of f_0 is enhanced during perturbations that oppose the direction of the planned change in f_0 ; however, the results indicate that task differences were only observed in the shorter latency of peak compensatory responses to upward perturbations, contrary to the hypothesis.

There are several possibilities as to why results from the present study seemingly contradict those of [Chen et al. \(2007\)](#). Although a common motor goal in both tasks is to raise f_0 during vocalization, the speech task in the study by [Chen et al. \(2007\)](#) contains linguistic and prosodic intent which may compound to form a stronger intention with the motor intent. In speech tasks, the overall vocal pitch changes are a means by which linguistic or prosodic intent is achieved, whereas in the current study, the change in f_0 is made irrespective of a linguistic goal and hence may be considered the primary goal. In response to upward perturbations, compensatory responses would have caused voice f_0 to decrease from baseline, which conflicts with the intended plan to raise f_0 . Purely in terms of dynamic f_0 control, the auditory-vocal system may be sensitive to mismatches in the estimated state of the vocal apparatus with the intended motoric changes, leading to a shorter peak compensation latency in order to smoothly execute the increase in f_0 ([Houde and Nagarajan, 2011](#)).

Likewise, the timing of the perturbation in the study by [Chen et al. \(2007\)](#) may have occurred at a different time in relation to the f_0 increase when compared to the present study. The previous study ([Chen et al., 2007](#)) used an audio model for their speech task but did not offer further control in regards to timing during the actual production, potentially affecting when the perturbation may have fallen during the phrase. The present study attempted to build stronger control over these parameters by using visual aids that guided the participant throughout the course of a planned f_0 change. These differences in timing may have captured different stages of the vocal motor planning process, particularly with the perturbation timing in the present study occurring earlier, prior to implementing any changes in vocal motor control. Furthermore, different stimulus parameters between the two studies may have also contributed to the contrasting results. Compensatory responses in the speech task ([Chen et al., 2007](#)) were accessed using three magnitudes of perturbation (± 50 , ± 100 , ± 200). It is important to note that although

Chen *et al.* (2007) reported sensitivity to downward perturbations in the speech task, they also reported an increase in compensatory response magnitude with increased stimulus magnitude. The results appear to trend toward a three way interaction in which the sensitivity to downward perturbations in the speech task appear to be largely due to the responses to -200 cent perturbations (Chen *et al.*, 2007). Liu *et al.* (2009) likewise acknowledge that despite the similarities between the speech tasks in their study and the study by Chen *et al.* (2007), they also did not observe differences in response magnitudes to any particular stimulus direction, possibly because their stimulus magnitudes (± 100 cents) were not large enough. Taken together, these results suggest that the relationship between feedback and the intended f_0 change may be more related to a mismatch between the current f_0 state, depending on the goal of the vocal task, and that directional sensitivity to perturbation directions may be apparent when there is a substantial mismatch.

Although the present study originally planned to compare the raised f_0 task to a decreased f_0 task, the majority of participants produced glottal fry when attempting to lower their voice f_0 . Therefore, the data was not analyzable for the purposes of this study. A potential future direction of this research could be to provide greater experimental control for the recording of a decreased f_0 task. The explanations provided in this present study would be strengthened if a planned decrease in f_0 would be accompanied by shorter latencies of compensatory responses to downward perturbations (-50 cents), providing a counterpart for the shorter latencies of compensatory responses to upward perturbations ($+50$ cents) in raised f_0 tasks.

C. Modulatory effects of musicality on compensatory and voluntary f_0 control

Previous studies have reported on the differences in compensatory auditory-vocal control between non-musicians and musicians (Keough and Jones, 2009; Zarate and Zatorre, 2008; Behroozmand *et al.*, 2014; Sturgeon *et al.*, 2015). A comparison between the non-musician and musician groups in the current study showed that musicians produced larger compensatory responses in steady f_0 tasks, which is in agreement with the literature on small magnitude, short-duration pitch-shifts (Behroozmand *et al.*, 2014; Sturgeon *et al.*, 2015). The present study also demonstrated that musically-trained individuals produced larger compensatory responses prior to planned changes in f_0 , compared to non-musicians. These results suggest that musical training provides general changes in auditory-motor integration that is more sensitive to errors in feedback. Although pitch perturbations were randomized across stimulus direction and vocal task while also including control trials, they were always applied at a fixed time (1200 ms after voice onset). This feature of the pitch-shift parameters may have inadvertently caused participants to form expectancies regarding the perturbation timing (Korzyukov *et al.*, 2012) and potentially masked any task differences. Future studies should attempt to randomize the pitch-shift onset to further explore the sensitivity of unanticipated feedback errors across tasks and musicality.

Interestingly, the presence of perturbations appears to have affected how participants adjusted their forthcoming voluntary f_0 increases. Although musicians on average still produced peak voluntary f_0 changes with shorter latencies, they were significantly slower in perturbed trials compared to control trials. Inversely, non-musicians had shorter latencies in their peak voluntary f_0 changes for perturbed trials compared to control trials. A possible explanation for the reversal in behavior may be due to the difference in how individuals with musical training incorporate auditory feedback during vocal motor control. When asked to volitionally ignore or compensate for perturbations in pitch auditory feedback, the neural activity in trained singers, compared to that of non-musicians, has been shown to recruit areas such as the auditory cortices, anterior cingulate cortex, and putamen which allow for more nuanced auditory feedback monitoring and pitch control (Zarate and Zatorre, 2008). The introduction of pitch-shifts during a volitional task may have required additional auditory processing for musicians in order to incorporate the deviations in feedback. Therefore, in the present study, musicians may have had to reduce the speed of their vocal motor movements in order to maintain the accuracy of their intended f_0 increase while incorporating corrections for the feedback error. In contrast, the study by Zarate and Zatorre (2008) also demonstrated that whereas the trained singers were able to volitionally suppress their compensatory responses and maintain their feedforward control of a steady vowel vocalization, non-musicians were unable to suppress their compensatory responses. Similarly, in the present study, the non-musicians may not have been able to integrate their auditory feedback and vocal motor control as efficiently, and the automatic nature of the pitch-shift response may have disrupted and influenced the timing of their feedforward control, such that they initiated their planned increase in f_0 earlier than intended. These converging lines of research indicate that musicians are able to more robustly compensate for mismatches between sensory expectations and feedback while simultaneously being able to regulate their vocal f_0 control so as to not veer away from their intended motor plans. On the other hand, non-musicians may not be able to achieve the balance between sensorimotor feedback and ongoing volitional control as efficiently as musicians.

However, future studies should also make greater attempts to reduce the variability amongst participants in regards to the peak voluntary f_0 change. Although the magnitude of the peak voluntary f_0 change was not statistically significant between groups, both between-group and within-group variability was large, potentially masking some effects. A potential solution to this problem could be to include a pitch target that participants can raise their f_0 towards from a steady, conversational f_0 level or to include a larger sample size for both groups.

V. CONCLUSION

The present study investigated the roles of musicality and planning on auditory-vocal feedback integration during volitional voice f_0 control. Prior to raising voice f_0 from a steady, conversational level, participants received and involuntarily compensated for unexpected pitch perturbations of ± 50 cents. Although compensation magnitudes were larger overall in

response to -50 cent stimuli, participants had shorter latencies in peak compensatory response for $+50$ cent stimuli when they were planning to increase their f_0 immediately following the perturbation, compared to trials where they kept their f_0 at a steady, conversational level. In contrast, previous studies using speech tasks found that compensatory responses were more sensitive when the error in pitch auditory feedback conflicted with the direction of an intended pitch inflection. These results suggest that during speech tasks, speakers may respond more selectively to perceptual mismatches with intended suprasegmental features, while dynamic f_0 tasks may use state-estimated feedback control to mediate between the current dynamic state and planned motor changes. Across both tasks, musicians produced larger compensatory response magnitudes compared to non-musicians. Peak voluntary f_0 change latencies also indicate that musicians are more efficient in feedforward f_0 control compared to non-musicians on average, but an inverse in the voluntary change latency during perturbed trials indicate that the level of musicality may change the strategies used to incorporate auditory feedback into voluntary f_0 control. Overall, the results presented in this study suggest that the auditory-vocal system is contextually sensitive to pitch errors during feedforward planning and that musical training affects how efficiently individuals can integrate sensorimotor adjustments during dynamic voice control.

ACKNOWLEDGMENTS

This study was supported by a grant from the NIH, Grant No. 1R01DC006243. The authors would like to thank C.L. Chan for his help with the computer programming.

- Bauer, J. J., and Larson, C. R. (2003). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," *J. Acoust. Soc. Am.* **114**, 1048–1054.
- Bauer, J. J., Mitall, J., Larson, C. R., and Hain, T. C. (2006). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," *J. Acoust. Soc. Am.* **119**(4), 2363–2371.
- Bernsee, S. M. (1999). "Routine for doing pitch shifting while maintaining duration using the Short Time Fourier Transform, [computer program]," Version 1.2, retrieved from <http://blogs.zynaptiq.com/bernsee/pitch-shifting-using-the-ft/> (Last viewed October 29, 2018).
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am.* **103**, 3153–3161.
- Burnett, T. A., and Larson, C. R. (2002). "Early pitch shift response is active in both steady and dynamic voice pitch control," *J. Acoust. Soc. Am.* **112**, 1058–1063.
- Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., and Larson, C. R. (2014). "Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch," *Brain Cogn.* **84**(1), 97–108.
- Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C. R. (2012). "Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control," *J. Acoust. Soc. Am.* **132**, 2468–2477.
- Boersma, P., and Weenink, D. (2017). "Praat: Doing phonetics by computer, [computer program]," Version 6.0.29, <http://www.fon.hum.uva.nl/praat/> (Last viewed July 3, 2017).
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice F0 responses to pitch-shifted voice feedback during English speech," *J. Acoust. Soc. Am.* **121**, 1157–1163.
- Donath, T. M., Natke, U., and Kalveram, K. T. (2002). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," *J. Acoust. Soc. Am.* **111**, 357–366.
- Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., and Eisner, F. (2018). "Opposing and following responses in sensorimotor speech control: Why responses go both ways," *Psychonom. Bull. Rev.* **25**(4), 1458–1467.
- Hain, T. C., Burnett, T. A., Larson, C. R., Singh, S., and Kenney, M. K. (2000). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," *Exp. Brain Res.* **130**, 133–141.
- Halwani, G. F., Loui, P., Rüber, T., and Schlaug, G. (2011). "Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians," *Front. Psychol.* **2**(156), 1–9.
- Houde, J. F., and Nagarajan, S. S. (2011). "Speech production as state feedback control," *Front. Hum. Neurosci.* **5**(82), 1–14.
- Jones, J. A., and Munhall, K. G. (2002). "The role of auditory feedback during phonation: Studies of Mandarin tone production," *J. Phonetics* **30**, 303–320.
- Keough, D., and Jones, J. A. (2009). "The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing," *J. Acoust. Soc. Am.* **126**, 837–846.
- Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., and Lotze, M. (2010). "The brain of opera singers: Experience-dependent changes in functional activation," *Cereb. Cortex* **20**(5), 1144–1152.
- Korzyukov, O., Sattler, L., Behroozmand, R., and Larson, C. R. (2012). "Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback," *PLoS One* **7**(7), e41216.
- Korzyukov, O., Tapaskar, N., Pflieger, M. E., Behroozmand, R., Lodhavia, A., Patel, S., Robin, D. A., and Larson, C. R. (2015). "Event related potentials study of aberrations in voice control mechanisms in adults with attention deficit hyperactivity disorder," *Clin. Neurophysiol.* **126**(6), 1159–1170.
- Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (2000). "Effects of pitch-shift onset velocity on voice F_0 responses," *J. Acoust. Soc. Am.* **107**, 559–564.
- Larson, C. R., and Robin, D. A. (2016). "Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control," *AIMS Neurosci.* **3**(1), 22–39.
- Liu, H., and Larson, C. R. (2007). "Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex," *J. Acoust. Soc. Am.* **122**, 3671–3677.
- Liu, H., Meshman, M., Behroozmand, R., and Larson, C. R., (2011). "Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback," *Clin. Neurophysiol.* **122**(5), 951–957.
- Liu, H., Xu, Y., and Larson, C. R. (2009). "Attenuation of vocal responses to pitch perturbations during Mandarin speech," *J. Acoust. Soc. Am.* **125**, 2299–2306.
- Nardo, D., and Reiterer, S. M. (2009). "Musicality and phonetic language aptitude," in *Language Talent and Brain Activity*, edited by G. Dogil and S. M. Reiterer (Mouton de Gruyter, Berlin), pp. 213–256.
- Natke, U., Donath, T. M., and Kalveram, K. T. (2003). "Control of voice fundamental frequency in speaking versus singing," *J. Acoust. Soc. Am.* **113**, 1587–1593.
- Natke, U., and Kalveram, K. T. (2001). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," *J. Speech Lang. Hear. Res.* **44**, 577–584.
- Parkinson, A. L., Behroozmand, R., Ibrahim, N., Korzyukov, O., Larson, C. R., and Robin, D. A. (2014). "Effective connectivity associated with auditory error detection in musicians with absolute pitch," *Front. Neurosci.* **8**(46), 1–9.
- Patel, S., Nishimura, C., Lodhavia, A., Korzyukov, O., Parkinson, A., Robin, D. A., and Larson, C. R. (2014). "Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback," *J. Acoust. Soc. Am.* **135**(5), 3036–3044.
- Scheerer, N. E., and Jones, J. A. (2018). "Detecting our own vocal errors: An event-related study of the thresholds for perceive and compensating for vocal pitch errors," *Neuropsychologia* **114**, 158–167.
- Sturgeon, B. A., Hubbard, R. J., Schmidt, S. A., and Loucks, T. M. (2015). "High F_0 and musicianship make a difference: Pitch-shift responses across the vocal range," *J. Phon.* **51**, 70–81.
- Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (2004). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," *J. Acoust. Soc. Am.* **116**, 1168–1178.
- Zarate, J. M. (2013). "The neural control of singing," *Front. Hum. Neurosci.* **7**(237), 1–12.
- Zarate, J. M., Wood, S., and Zatorre, R. J. (2010). "Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers," *Neuropsychologia* **48**(2), 607–618.
- Zarate, J. M., and Zatorre, R. J. (2008). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," *Neuroimage* **40**(4), 1871–1887.