Comparison of volitional opposing and following responses across speakers with different vocal histories

Sona Patel,[1,a)] Li Gao,[2] Sophie Wang,[2] Christine Gou,[2] Jordan Manes,[2] Donald A. Robin,[3] and Charles R. Larson[2]

[1]Department of Speech-Language Pathology, Seton Hall University, 340 Kingsland Street, Building 123, Nutley, New Jersey 07110, USA

[2]Northwestern University, Frances Searle Building, 2240 Campus Drive, Room 3-247, Evanston, Illinois 60208-2952, USA

[3]Communication Sciences and Disorders, University of New Hampshire, Hewitt Hall, Room 153, Durham, New Hampshire 03824, USA

Research has shown that people who are instructed to volitionally respond to pitch-shifted feedback either produce responses that follow the shift direction with a short latency of 100–200 ms or oppose the shift direction with longer latencies of 300–400 ms. This difference in response latencies prompted a comparison of three groups of vocalists with differing abilities, non-trained English-speaking subjects, non-trained Mandarin-speaking subjects, and trained English-speaking singers. All subjects produced short latency following responses and long latency opposing responses, and in most cases the opposing responses were preceded by a shorter latency following response. Across groups, the magnitudes of the opposing and following responses were largest for the Mandarin speakers. Singers produced the smallest opposing response magnitudes, suggesting differences in the pitch goals of the two groups. Opposing response latencies were longest for the English and Mandarin speaking subjects and shortest for the trained singers, demonstrating that musical training increases the speed of producing the opposing responses. The presence of similar latencies of small following responses preceding larger opposing responses in all groups suggests that the tendency to mimic changes in sounds to which a person is attending are not influenced by vocal training or experience.
© 2019 Acoustical Society of America. https://doi.org/10.1121/1.5134769

I. INTRODUCTION

During the last 15 years, considerable progress has been made in understanding the role of voice auditory feedback in the neural control of voice fundamental frequency ($F0$) and intensity (Bauer *et al.*, 2006; Behroozmand *et al.*, 2012; Burnett *et al.*, 1998; Chen *et al.*, 2007; Donath *et al.*, 2002; Hafke, 2008; Hain *et al.*, 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001; Patel *et al.*, 2014). These studies showed that when subjects vocalized and heard an unexpected change in their voice pitch or loudness, they produced an automatic, reflexive, response that changed their voice $F0$ or intensity in the opposite direction of the auditory stimulus, as if correcting for an error in production (i.e., an "opposing" response). Though less frequent, there were also automatic responses that changed in the same direction as the stimulus, described as "following" responses. Both the opposing and following responses had latencies of about 200 ms. Because of their direction, the opposing responses were suggestive of a negative feedback system that functioned to correct for errors in voice $F0$ production.

In contrast with the above experiments that reflected automatic mechanisms of voice control, in a new paradigm (Patel *et al.*, 2014) subjects were asked to volitionally change their voice $F0$ when they heard a change in the pitch of their voice auditory feedback. It was found that when subjects produced volitional responses that either opposed or followed the direction of a pitch-shift, the following responses had a significantly shorter latency (<200 ms) than the opposing responses (>300 ms). Another interesting finding was the presence of a small automatic following response prior to the opposing response production when instructed to produce an opposing response (Patel *et al.*, 2014). Thus, in both of the above conditions, the following responses had a much shorter latency than the opposing responses. All of these observations suggest that neural mechanisms of vocal control based on auditory feedback are more varied than previously recognized and are highly dependent on whether or not a subject intends to react to a change in voice pitch auditory feedback.

Another method used to study voice control is to require subjects to shadow a "side-tone" (i.e., a tone that accompanies a vocalist) by changing the pitch of their voice to follow the direction of the side-tone (Bailly, 2003; Horii, 1979; Leonard *et al.*, 1988; Leonard and Ringel, 1979; Peschke *et al.*, 2009; Shockley *et al.*, 2004). The reported shadowing response latencies were less than 200 ms, which are similar to the volitional following responses found by Patel and colleagues (2014) and the involuntary following response latencies obtained with pitch-shifted auditory feedback

a)Electronic mail: Sona.patel@shu.edu

[e.g., Burnett *et al.* (1998)]. Also, the volitional nature of the procedures in the shadowing experiments and the following responses in the study by Patel *et al.* (2014) suggest these two types of volitional responses may involve the same or similar neural mechanisms. The mechanisms for these types of responses are unknown, but may reflect a "feedforward" process, i.e., a process for changing voice $F0$ without the need for guidance by sensory feedback. Patel *et al.* (2014) used untrained, native English speaking subjects to study voluntary following and opposing responses to feedback perturbations. In studies of involuntary responses to pitch-shifted feedback, it was found that vocal training or language background affects the magnitude and speed of the response. For example, trained singers produced faster voice $F0$ responses (i.e., had smaller latencies) than non-trained singers (Behroozmand *et al.*, 2014). Liu *et al.* (2010a) demonstrated that response magnitudes to pitch-shifted feedback were larger during a sustained high frequency $F0$ compared to low frequency $F0$. Liu *et al.* (2010b) demonstrated that English-speaking children produced smaller response magnitudes than adults, while Mandarin-speaking children produced larger response magnitudes than adults. Ning *et al.* (2015) found that Mandarin speakers had a smaller response magnitude, i.e., better pitch control, compared to English speakers and singers under both bi-tonal syllable production (/ma/) and single vowel production.

These data on involuntary responses suggest that the neural mechanisms of voice control vary as a function of vocal use or training. It is unknown whether training or language background would similarly affect *volitional* responses to pitch-shifted feedback. Volitional responses are clearly affected by factors such as age (Nagao *et al.*, 2008) and may be susceptible to musical training and language backgrounds. Only a few studies have examined the role of "pitch training" on responses to pitch-shifted feedback during dynamic voice $F0$ changes. For example, Xu *et al.* (2004) examined responses to pitch-shifts during bi-tone sequences representative of Mandarin speech. Results showed that speakers of a tonal language (Mandarin) produced faster (~50 ms) and larger (~35 cents) responses than speakers of a non-tonal language (English). These results examine shifts during a dynamic motor target. Critically, if volitional responses made in response to an unexpected change in voice pitch auditory feedback vary across different vocal backgrounds or abilities, then we can hypothesize that the neural circuitry underlying these responses can be altered by a variety of contexts.

In the present study, we tested three groups of speakers varying in the way they use pitch: a group of young adults who were trained singers and spoke English as their native language ("Singers"), a group of untrained English speaking subjects ("English"), and a group of individuals who spoke Mandarin as their native language ("Mandarin"). Participants were asked to volitionally and dynamically change their pitch in response to a perceived change in their voice pitch (a pitch-shift stimulus). It was hypothesized that vocal training and the use of pitch in a tonal language (dramatically increasing the need for precise vocal control compared to non-tonal speakers) will change the latency or magnitude of volitional responses made due to an unexpected change in voice pitch

auditory feedback. It is well known that speakers of tonal languages engage the right hemisphere for speech production more than speakers of non-tonal languages (Chandrasekaran *et al.*, 2007; Chen *et al.*, 2012; Zatorre and Gandour, 2008). In addition, well-trained musicians rely on right hemisphere mechanisms for vocal control to a greater extent than non-trained musicians (Behroozmand *et al.*, 2014; Takeuchi and Hulse, 1993; Zatorre, 2001, 2003). It is interesting to consider that voice pitch also seems to be right lateralized (Belin, 2006; Belin and Zatorre, 2003; Gandour *et al.*, 2004; Tourville *et al.*, 2008). We hypothesize that engagement of right hemisphere mechanisms for tonal language speakers and singers is due to their complex use of pitch, and as a result of their additional experience with intentional pitch manipulation, we predicted that both Mandarin speakers and singers would produce more rapid volitional changes in voice $F0$ than the English speakers. This prediction is supported by previous findings in studies of involuntary responses to pitch-shifts in singers (Keough and Jones, 2009) and Mandarin speakers (Xu *et al.*, 2004).

One factor to consider is that trained musicians spend years practicing to produce specific musical notes with great accuracy and to respond accurately to changes in auditory feedback of musical notes, such as from other singers or a musical instrument. Jones and Keough (2008) report that singers may rely more on an internal model for $F0$ production during singing compared to non-singers. The use of pitch to produce a phonemic distinction is very different from its use in singing. Two particularly important differences are: (1) the linguistic use of pitch is always produced with a very short duration (the time of a phoneme) whereas singers modulate pitch across entire melody lines, and (2) the use of pitch to communicate when singing occurs through suprasegmental features (as in prosody) instead of phonology. Therefore, a second hypothesis is that individuals trained in the linguistic use of pitch (Mandarin group) may differ in response magnitude from individuals trained in using pitch for melody (Singers group). Specifically, we predict, based on the results of Keough and Jones (2009) that singers' may attempt to match the magnitude of the pitch-shift even though not explicitly instructed to do so. This may result in a response curve that quickly reaches an intended pitch target that is maintained for the remainder of the vocalization. For this reason, the absolute value of the peak response magnitude may be smaller than Mandarin and English speakers, who may continue to increase or decrease their pitch rather than reach any specific pitch target.

II. METHODS

A. Participants

Three groups of healthy young adults with no reported neurological or hearing disability participated in this study. The English group consisted of 13 subjects (10 females, age: 20–30 years). These subjects reported no formal musical training and spoke English as their primary language. The Mandarin group consisted of 14 native Chinese subjects (11 females, age: 19–33 years) who were bilingual in English. These individuals had no singing training. The trained singer group consisted of 10 subjects (7 females, age: 18–28 years)

recruited from the Northwestern University Bienen School of Music. All subjects were healthy at the time of testing and were not experiencing upper or lower respiratory infections. All subjects signed informed consent approved by the Northwestern University Institutional Review Board.

B. Procedures

All testing was done with subjects seated in a sound-attenuated room. A microphone (AKG boomset microphone, model C420, Harmon International Industries Inc., Stamford, CT) was placed 1 in. in front of the mouth to record the voice as subjects vocalized an /a/ vowel. Subjects' voices were amplified with a Mackie mixer (model 1202-VLZ3, Loud Technologies, Woodinville, WA), pitch-shifted through an Eventide Eclipse Harmonizer (Eventide, Inc., Little Ferry, NJ), and then presented to the ears via headphones (Sennheiser) using musical instrument digital interface software (Max/MSP v.5.0 Cycling '74, Walnut, CA). This feedback signal (i.e., the pitch-shifted version of the participant's voice) was presented at about 80–85 dB. The 10-dB gain in the feedback channel relative to vocal output (controlled by a Crown Audio Inc. D75 amplifier, Elkhart, IN) was used to partially mask air- and bone-conducted voice feedback. The pitch-shift stimuli had a rise and fall time of approximately 15 ms, which is fast enough that subjects cannot detect the rising or falling change in voice pitch feedback. A transistor-transistor logic (TTL) pulse synchronized with the presentation of the stimulus was recorded along with the subject's voice and the voice feedback signal using Chart software on a Powerlab A/D convertor (model 880, AD Instruments, Castle Hill, Australia) at a sampling frequency of 10 kHz. Following testing, data were analyzed using procedures written for Igor Pro (v6.0, Wavemetrics Inc., Lake Osewago, OR).

Subjects were seated in front of a monitor. Before beginning the test, subjects were told that the monitor would display instructions such as "Say aah" or "Wait" and that a progress bar will indicate when they should vocalize and the duration of their vocalization. They were also told that as they vocalize, the sound of their voice will be different from what they expect. Specifically, their voice pitch will be higher or lower than expected. Then they were told that their job is to change the pitch of their voice in one of two ways, corresponding to the two conditions in this study. In the opposing condition, subjects were instructed to change the pitch of their voice in the opposite direction to any pitch changes in the auditory feedback heard through the headphones. That is, if they heard their voice feedback increase in pitch, they should lower their pitch. In the following condition, subjects were instructed to change the pitch of their voice to match that of a change in pitch of the voice auditory feedback signal. Thus, if they heard their voice pitch feedback increase in pitch, they should increase the pitch of their voice. Subjects were not instructed how fast to change the pitch of their voice or how much to change it, only that they should be accurate in the direction of their change in voice $F0$. Given that we predicted changes in timing, we instructed participants to increase or decrease their pitch in the appropriate direction and then hold that pitch until the end of the

trial. In both conditions, the subjects vocalized for about 2.5 s in each of 4 blocks of 52 vocalizations. Pitch shifts were presented at a random time (0.5 to 1 s) after vocal onset, in the amount of either 100 cents up or 100 cents down for 1 s.

C. Analyses

For voice analysis, $F0$ contours were first generated using Praat (Boersma and Weenink, 2001). TTL pulses were used to align the vocal signals with the onset of the pitch-shift stimulus using Igor Pro software. Then the $F0$ contours were converted to cents using the formula in Eq. (1),

$$1200 * \log_2(f_2/f_1). \tag{1}$$

The f_1 value represents the reference frequency measured every 5 ms in the pre-shift window (200 ms prior to the shift), and f_2 represents the $f0$ value measured in the post-stimulus time window every 5 ms. Then, all pitch segments (-400 before to 1000 ms after stimulus) of responses that either followed or opposed the stimulus direction were separately averaged for each subject. Igor scripts were used (Behroozmand *et al.*, 2012) in which individual responses for each stimulus were first isolated as to their direction, i.e., either in the opposite direction (opposing) or the same direction (following) as the stimulus. Trials that were in the wrong (unintended) direction were removed (approximately less than 5%–10% of trials). These were treated as errors and discarded. Further research is needed to understand why these occur, but this is outside the scope of the current paper. Thus, we only averaged congruent responses by task (e.g., opposing in opposing condition). Averaged response peak magnitudes were separately measured for both the opposing and following responses at the time of the greatest (increases in voice pitch) or lowest (decreases in voice pitch) value of the averaged response occurring after the stimulus onset. Averaged response onset times (latency) were measured for each subject at the time when the averaged curve first deviated from the pre-stimulus baseline by a value of 2 standard deviations (SDs) of the baseline average.

From the individual averaged vocal responses for each subject, grand averaged responses were calculated across all subjects by group. Averaged response peak magnitudes and latencies were tested with a repeated measures analysis of variance (RM-ANOVA) in SPSS, Inc. (Chicago, IL). We used within and between subjects comparisons that tested for differences in response magnitudes and latencies by condition (opposing vs following), perturbation direction (upwards or downwards), and group (English, Mandarin, and Singers). Follow-up independent sample *t*-tests were performed. An early response was discovered in the opposing condition for all groups. The mean latency of these responses was compared to the main opposing response latency and the main following response latency across groups using a RM-ANOVA and follow-up one-way ANOVAs and paired sample *t*-tests.

III. RESULTS

All groups of subjects responded with voice $F0$ changes in the correct direction as instructed. Figures 1, 2, and 3
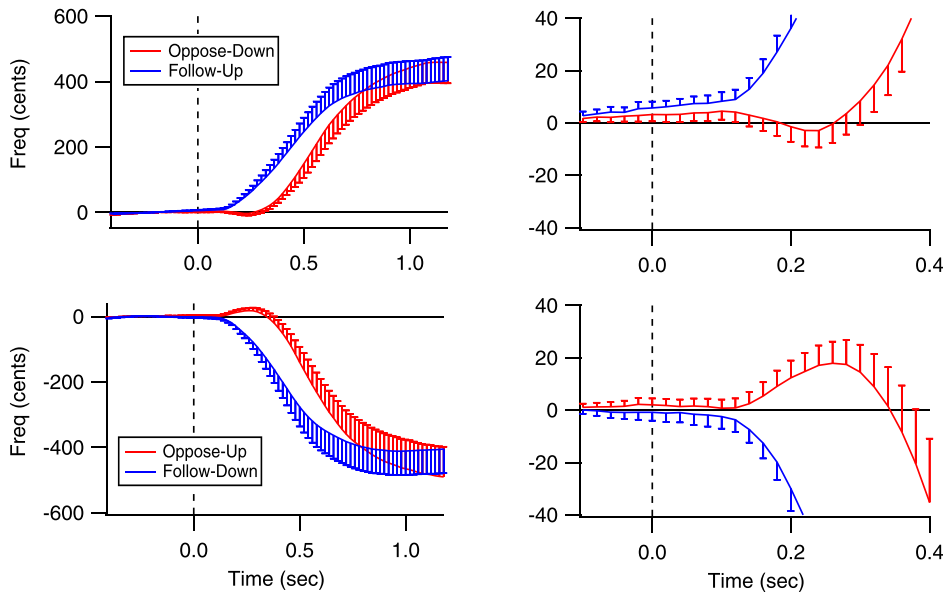
MANDARIN

FIG. 1. (Color online) Voice $F0$ contours for the opposing and following responses produced by the Mandarin speakers. The vertical dashed line indicates the time at which the pitch-shift stimulus occurred (time $= 0$ s). The blue traces represent following responses (top row: to an upward shift; bottom row: to a downward shift) in voice pitch feedback, and the red traces represent opposing responses to a downward pitch shift (top row: to a downward shift; bottom row: to an upward shift). Error bars indicate standard error of the mean $F0$. The plots on the right show an expanded portion of the upper traces to illustrate that the opposing responses are preceded by a small tendency to follow the stimulus direction.

illustrate grand average responses for all groups of subjects. In each figure, the voice responses are shown with error bars indicating standard error of the mean $F0$. Upward responses are shown in the top row and downward responses in the bottom row. The plots in the left column are plotted with magnitude and time scales to illustrate the entire responses to 1200 ms post stimulus onset. The plots in the right column have a smaller magnitude and temporal scale to illustrate details of the earliest changes in the vocal responses that are not easily seen in the plots on the left. The plots on the left show that when following the stimulus direction (blue traces), responses are faster than when opposing the stimulus direction (red traces). Also, response magnitudes were larger than 100 cents, the pitch-shifted value. Although this is likely impossible for involuntary responses, the present study examined volitional responses in which through pilot

testing it was observed to be difficult to make smaller pitch changes than larger ones. The large magnitudes are likely due to the open-ended instructions to participants regarding how high or low change their pitch, resulting in exaggerated changes in pitch. Another observation is that for each subject group and each stimulus direction, the responses to the following condition began earlier than the responses to the opposing condition. Across all groups, the average response magnitude of the opposing condition was 376 cents (SD: 215.9) and in the following condition, 357 cents (SD: 190).

For measures of response magnitude, the absolute values of the responses were analyzed to account for the downward responses that had negative values. The RM-ANOVA of response magnitudes revealed a significant main effect for group [$F(2,39) = 7.514$, $p < 0.01$]. A significant interaction of group by perturbation direction by condition [$F(2,39)$
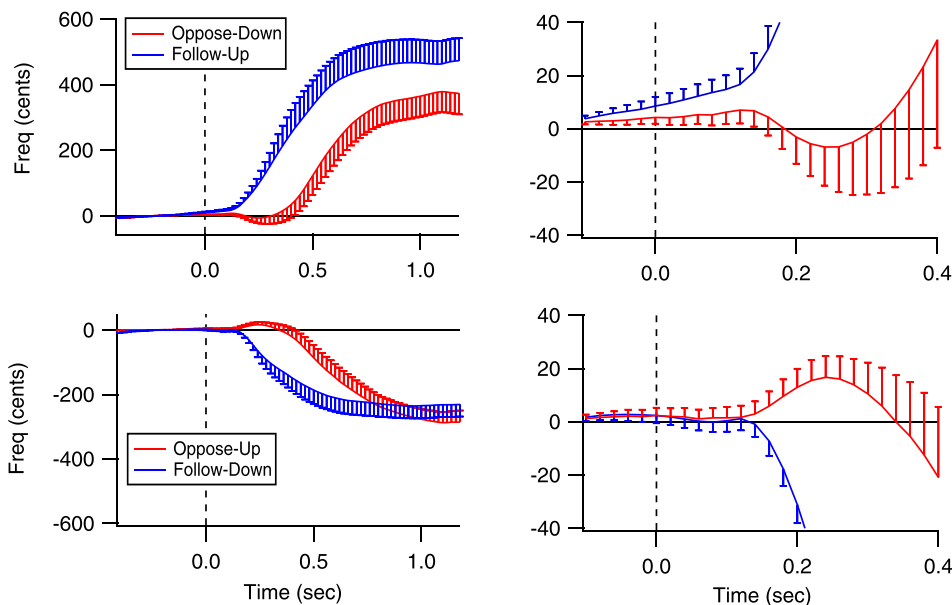
ENGLISH

FIG. 2. (Color online) Voice $F0$ contours for the English speakers. The vertical dashed line indicates the time at which the pitch-shift stimulus occurred (time $= 0$ s). The blue traces represent following responses (top row: to an upward shift; bottom row: to a downward shift) in voice pitch feedback, and the red traces represent opposing responses to a downward pitch shift (top row: to a downward shift; bottom row: to an upward shift). Error bars indicate standard error of the mean $F0$. The plots on the right show an expanded portion of the upper traces to illustrate that the opposing responses are preceded by a small tendency to follow the stimulus direction.

J. Acoust. Soc. Am. **146** (6), December 2019

Patel *et al.*   4247
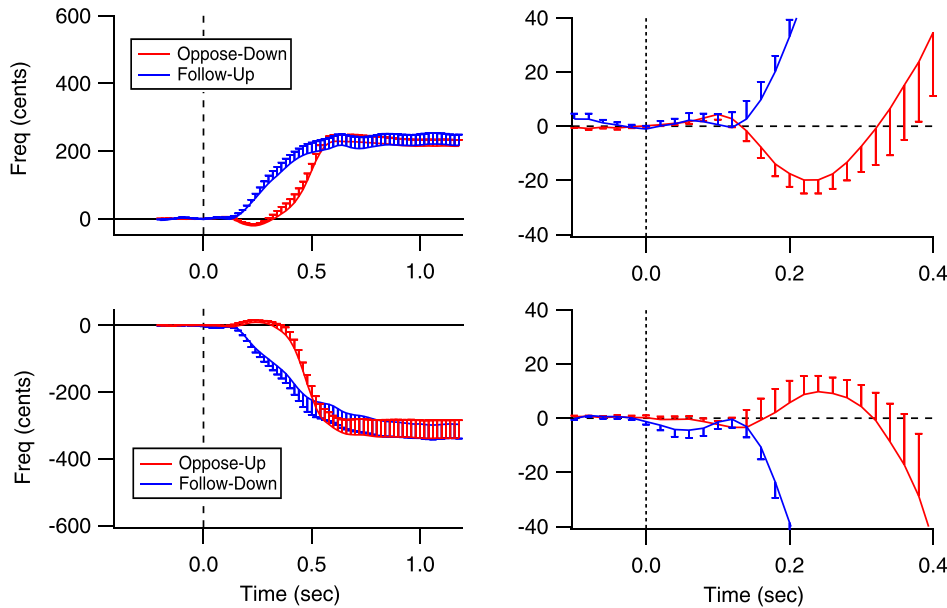
TRAINED SINGERS

FIG. 3. (Color online) Voice $F0$ contours for the trained singers. The vertical dashed line indicates the time at which the pitch-shift stimulus occurred (time = 0 s). The blue traces represent following responses (top row: to an upward shift; bottom row: to a downward shift) in voice pitch feedback, and the red traces represent opposing responses to a downward pitch shift (top row: to a downward shift; bottom row: to an upward shift). Error bars indicate standard error of the mean $F0$. The plots on the right show an expanded portion of the upper traces to illustrate that the opposing responses are preceded by a small tendency to follow the stimulus direction.

= 5.209, $p < 0.01$] was also found. No other interactions were significant. The mean response magnitude was largest for the Mandarin speakers (M: 460.4 cents; SD: 254.9) compared with English speakers (M: 352.0 cents, SD: 183.0) and Singers (M: 283.3 cents; SD 112.9). Three follow-up independent sample t-tests comparing the magnitude of each group confirmed that Mandarin speakers' responses were larger than both English speakers [$t(110) = 2.585$, $p < 0.05$] and singers [$t(110) = 4.692$, $p < 0.05$], and that English speakers produced larger responses compared to singers [$t(110) = 2.291$, $p < 0.05$]. Table I displays the mean response magnitudes (and standard deviations) for each of the subject groups for opposing and following directed responses. Figures 4 and 5 show boxplots of response magnitude and latency for all subject groups in both the opposing and following conditions.

Table II displays the mean response latencies (and SDs) for each of the subject groups for opposing and following directed responses, and Fig. 6 illustrates relationships of the mean latencies for each condition and perturbation direction by group. RM-ANOVA of response latencies revealed a significant main effect for condition [$F(1,39) = 191.13.1$, $p < 0.01$] and perturbation direction [$F(1,39) = 4.853$, $p < 0.05$], but not for subject group [$F(2,39) = 2.161$, $p > 0.05$]. Significant interactions include direction by group [$F(2,39) = 3.674$, $p < 0.05$], condition by direction [$F(2,39) = 5.110$, $p < 0.05$], and condition by group by direction [$F(2,39) = 3.277$, $p < 0.05$]. For each subject group and perturbation direction, the mean latency of the following responses was on average 232 ms (range:

223–238 ms) shorter than the opposing responses. Voice latencies were shorter for downward perturbations in the opposing condition (i.e., upward responses) in Mandarin and English speakers, but the opposite occurred in Singers. Latencies in the following condition followed the opposite trend but the difference between upward and downward response times was small (an average of 7 ms compared to 99 ms in the opposing condition).

Another interesting feature of the responses is that for each group in the opposing condition, the majority of subjects (English, 10 subjects, Mandarin, 8 subjects, Singers, 6 subjects) produced an early response prior to the volitional opposing response. The early response was generally quite small and followed the direction of the pitch-shift stimulus (right column plots, in Figs. 1–3). A RM-ANOVA was performed for response latency (early, main opposing, main following) aggregated across direction, as not enough trials existed to look at effects of direction in a repeated design. Results showed main effects response latency [$F(2,60) = 116.118$, $p < 0.05$] but not group nor the group by latency interaction ($p > 0.05$). Follow-up one-way ANOVA showed that the mean latency of these early responses was similar across groups [English = 131 ms, Mandarin = 131 ms, Singers = 140 ms; $F(2,32) = 0.147$, $p > 0.05$]. Follow-up paired sample t-tests showed that these early following responses did in fact differ significantly from the main opposing response for each group [English: $t(11) = -6.032$, $p < 0.01$; Mandarin: $t(9) = -7.780$, $p < 0.01$; Singers: $t(10) = -11.921$, $p < 0.01$]. Paired sample t-tests also showed that these early responses

TABLE I. Mean response magnitudes in cents (and SDs) for each subject group and condition.

	Opposing Down	Opposing Up	Following Down	Following Up
Mandarin	485 cents (240)	525 cents (339)	461 cents (305)	396 cents (216)
English	373 cents (240)	234 cents (137)	256 cents (140)	436 cents (268)
Singer	266 cents (97)	333 cents (168)	314 cents (143)	220 cents (115)

Oppose Down

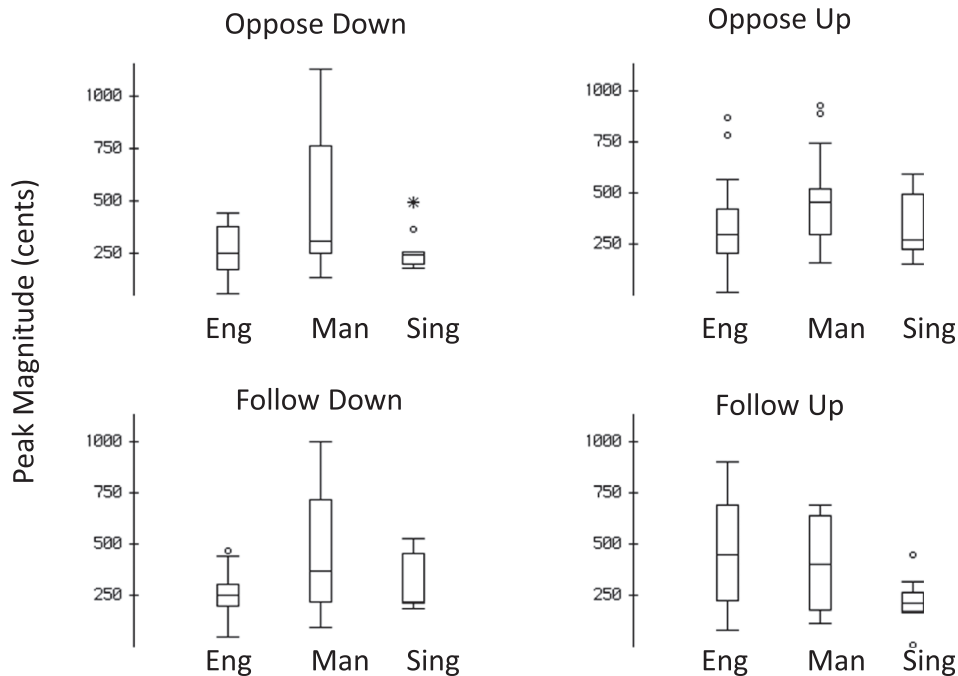## Oppose Up

Follow Down

Follow Up

FIG. 4. Boxplots of response magnitudes observed in the oppose-down, oppose-up, follow-down, and follow-up conditions. Boxplot definitions: Middle line is median, top, and bottom of boxes are the 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge–low hinge) and low hinge −1.5 (high hinge–low hinge).

were comparable to the response latencies for the following condition (English = 143 ms, Mandarin = 130 ms, Singers = 146 ms), for English speakers [$t(11) = -0.604$, $p > 0.05$], Mandarin speakers [$t(9) = 0.052$, $p > 0.05$], and Singers [$t(10) = 0.398$, $p > 0.05$].

IV. DISCUSSION

The present study was designed to determine if speakers with different vocal backgrounds and experiences (English speakers not vocally trained, Mandarin speakers, and English speakers trained in singing) react differently when instructed to volitionally change their $F0$ in a specific direction in response to perturbations in voice pitch auditory

feedback. Results show group differences in response magnitude and the presence of an "early response" in the opposing condition for all three groups. Below we specifically summarize findings pertaining to the response magnitude, response latency, and the presence of an early response.

A. Response magnitude

For the opposing responses, there were marked group differences in response magnitude. The Mandarin speakers produced the largest magnitude responses compared to the English-speaking subjects and the trained singers. This difference between Mandarin and English speakers may relate to the linguistic use of voice pitch in a tonal language. In

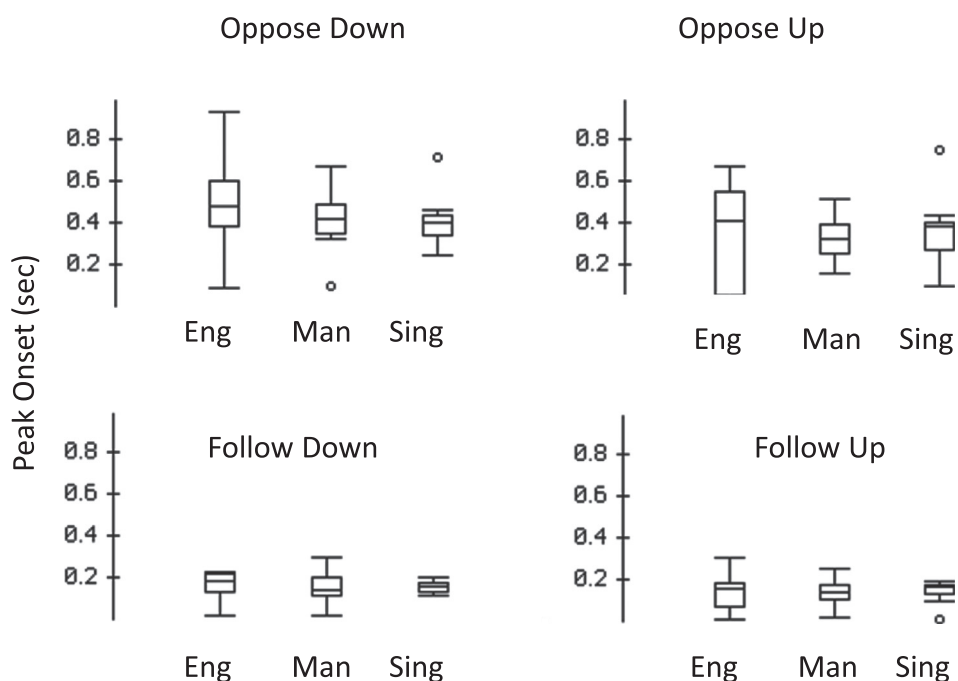## Oppose Down

Oppose Up

Follow Down

Follow Up

FIG. 5. Boxplots of response latencies observed in the oppose-down, oppose-up, follow-down, and follow-up conditions. Boxplot definitions: Middle line is median, top, and bottom of boxes are the 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge–low hinge) and low hinge −1.5 (high hinge–low hinge).

J. Acoust. Soc. Am. **146** (6), December 2019

Patel *et al.*   4249

TABLE II. Mean response latencies in ms (and SDs) for each subject group and condition.

	Opposing Down	Opposing Up	Following Down	Following Up
Mandarin	332 cents (113)	443 cents (97)	157 cents (51)	149 cents (47)
English	370 cents (217)	512 cents (154)	175 cents (34)	164 cents (62)
Singer	410 cents (106)	365 cents (143)	150 cents (23)	149 cents (24)

such languages, syllable-level changes in $F0$ are phonemically related to speech. It is possible that learning to produce such speech sounds, e.g., a falling tone, may lead to the ability to reduce $F0$ in a broader range of contexts, such as the volitional reductions in $F0$ observed in the present study. We believe that our results are not in contrast with others such as Ning *et al.* (2015), who found smaller response magnitudes (greater attenuation) during bi-tone production. Bi-tone production requires having specific pitch targets, and therefore, motor control can be demonstrated by the magnitude of the main response. Our task, on the other hand, was open-ended, with no clear pitch target instructed to the subject. The larger magnitude is probably more of an indicator of the scale size for the linguistic use of pitch.

While the singers generally produced the smallest responses, this was likely due to their singing training and resulting attempt to match their pitch target. These results are in line with Keough and Jones (2009), who reported that singers' $F0$ values were consistently closer to the intended pitch target in a study of involuntary responses to pitch shifts during the production of syllables in singers and non-singers. Singers reached a pitch asymptote at about 200–300 cents, which remained constant until the end of the analysis period (see Fig. 3). By contrast, the Mandarin and untrained English speakers' responses seemed to drift in magnitude until the end of the analysis period. The smaller variability in the singers' target pitch magnitude may reflect singing training, which involves learning a motor skill. The use of pitch in singing requires the precise use of the voice and exceptional vocal responses to auditory feedback (known to musicians as "the ear"). The steady response magnitudes produced by the singers probably relates to their better abilities to perceive and control voice pitch because of their extensive musical training.

Another difference between the three groups was the English speakers' difference in response magnitude for
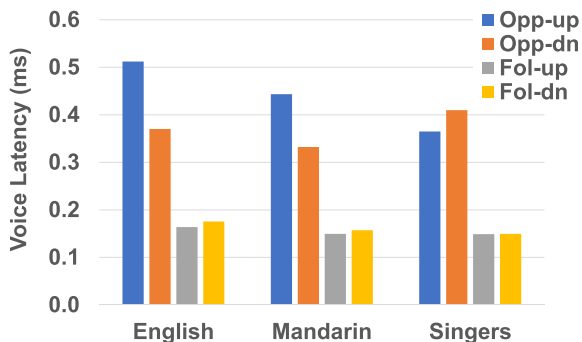

FIG. 6. (Color online) Mean latencies observed in the oppose-down, oppose-up, follow-down, and follow-up conditions for each group (English speakers, Mandarin speakers, and Singers).

downward responses. Mandarin speakers and singers produced responses in each response direction that were about equal in magnitude regardless of task. For example, in Mandarin speakers upward opposing responses were at the same magnitude as downward following responses (Figs. 1 and 3). The English speakers produced responses of differing magnitudes in the upward direction, specifically, following upward responses were much larger than opposing upwards responses (i.e., to downward shifts; see Fig. 2). Hain *et al.* (2000) performed a similar study in English speakers comparing pitch-shift responses when people were instructed to oppose and follow the shift among other tasks. They found no significant differences in response magnitude between opposing and following, however, their results were aggregated over up- and down-shift conditions. In our earlier work [Patel *et al.* (2014)], we also show no significant differences in opposing and following upward responses [p. 3039, Figs. 2(a) and 2(b)]. Thus, the reasons for this difference in upwards responses by task is not clear, although several studies have reported that shifts producing upward responses are more variable than downward shifts (Korzyukov *et al.*, 2015), possibly due to an inherent increased range of pitches available in the higher ends of our pitch range (Patel *et al.*, 2016).

B. Response latency

Contrary to our predictions, there were no significant differences in response latency between the groups. The wide variation in latencies in the opposing responses of the English subjects may have contributed to the failure to show significant results on these measures. Nevertheless, there were significant differences in response latencies between the opposing and following responses. The following response latencies were about 200 ms shorter than the opposing responses, which support results of a previous study (Patel *et al.*, 2014). The fact that there were no significant differences in following responses between subject groups suggests that these measures may reflect physiological limits to the speed at which humans can change the $F0$ of their voice in response to a change in a side-tone. Nevertheless, the high degree of variability in the opposing response latencies shown by the non-trained English speakers may reflect lack of training or the use of voice pitch linguistically, as seen in the Mandarin subjects.

C. Comparison of volitional and non-volitional vocal responses

Previous studies have mostly focused on examining the "basic pitch-shift response" or "pitch-shift reflex" (i.e., the "involuntary" response), which is the automatic and mostly compensatory pitch change obtained when people are asked to hold their voice steady regardless of changes to the pitch-altered auditory feedback. These studies (Bauer *et al.*, 2006; Behroozmand *et al.*, 2012; Burnett *et al.*, 1998; Chen *et al.*, 2007; Donath *et al.*, 2002; Hafke, 2008; Hain *et al.*, 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001) have shown that automatic (involuntary) compensatory responses have latencies in the range of 100–150 ms, which are similar to

the volitional following response latencies (130–150 ms) of the present study. Although the involuntary, pitch-shift reflex responses appear to function as a negative feedback system and the volitional following responses as a feedforward system, the similarities in latencies between these two conditions indicate they require the same processing times. The reason for the similarity of latencies in feedback in one task and feedforward in another may be the cognitive demands presented by both tasks. In the basic pitch-shift paradigm used to obtain pitch-shift reflex responses, participants are instructed to hold their voice constant. We would assume that such a task is not cognitively demanding ("basic pitch shift task" = low cognitive load). This is in contrast with the volitional paradigm, where participants are instructed to first listen for a pitch shift, and then change their voice pitch up or down based on the direction of the shift. Based on the number of steps alone, we argue that more cognitive processing is required for such a task (volitional task = high cognitive load).

On the other hand, the volitional opposing responses in the present study had latencies in the range of 350–500 ms, and importantly, these responses were preceded in most cases by small following responses. The observations of the small following responses that precede the delayed opposing responses suggest that when subjects are attending to a sound that changes in frequency, there is a strong tendency to match the change in frequency of the sound, which in turn may cause a delay in the onset of a volitional opposing vocal response. Vocal conditions in which these fast following responses might occur include vocal mimicry or shadowing.

The volitional following response latencies also were very similar to those reported in studies of vocal shadowing (Bailly, 2003; Horii, 1979; Leonard et al., 1988; Leonard and Ringel, 1979; Peschke et al., 2009; Shockley et al., 2004) or voice reaction time (Bakker and Brutten, 1989; Cross and Luper, 1983; Izdebski, 1980; Izdebski and Shipp, 1978; Shipp and Izdebski, 1975; Watson, 1994). Vocal mimicry has been observed in many animal species (Dalziell et al., 2015; Goodale and Kotagama, 2006; Kelley and Healy, 2010, 2011; Lilly, 1965; Reiss and McCowan, 1993; Richards et al., 1984) and is a highly conserved trait that may be inherited. Previous studies have suggested that vocal shadowing is similar to vocal mimicry and may reflect basic audio-motor transformations that are important for learning speech and language (Kuhl and Meltzoff, 1996; Wilbrecht and Nottebohm, 2003). These vocal abilities may also facilitate singers' ability to imitate another person's voice or to sing along with musical accompaniment.

One important question arising from the results of this study is regarding the underlying neural mechanisms controlling these responses. Although this study does not directly test the underlying neural mechanisms, Peschke et al. (2009) showed that vocal shadowing led to bilateral activation of the posterior and middle regions of the superior temporal gyrus, postcentral gyrus, inferior frontal gyrus, and precentral gyrus. These same neural locations were activated in response to pitch-shifted voice feedback (Parkinson et al., 2012; Peschke et al., 2009). Thus, it may be that variations in network connectivity may differentiate these two types of

processes rather than regional activation similarities. For example, we know that musicians with perfect pitch show different connectivity patterns than musicians with relative pitch or non-musicians (Parkinson et al., 2014). If in fact the following responses are a form of mimicry, it would be important to study them further in different groups such as children or those with neuromotor voice disorders. In regards to tonal languages, it would be important to learn if left hemisphere structures play a greater role than the right for the control of these responses given that changes in pitch are phonemic in nature.

Several studies have sought to describe involuntary responses to sensory stimulation from the modeling perspective. Functionally, control mechanisms of the involuntary opposing responses have been modeled as a negative feedback system (Behroozmand et al., 2009; Blakemore et al., 1999, 1998; Hain et al., 2000; Heinks-Maldonado and Houde, 2005; Houde and Jordan, 2002; Wolpert, 1997; Wolpert and Miall, 1996) or as a state feedback control system (Houde and Nagarajan, 2011). While there is merit for considering control of involuntary pitch-shift reflex responses from the perspective of system analyses, these approaches fail to explain the increase in response latencies for the volitional opposing responses in the present study.

One conceptualization of both the volitional following and opposing responses in the present and a previous (Patel et al., 2014) study is to consider an expanded model that places more emphasis on feedforward control (Lane et al., 2007; Wolpert, 1997). To explain both the following and opposing responses in the present study, Fig. 7 presents a model in which both feedforward and negative feedback control systems are operable. Given the speed of processing, it is likely that feedforward processing is the initial mode, in line with contemporary ideas of feedforward control (Ramanarayanan et al., 2016), and likely the default mode. Feedforward commands are learned over time and represent an average of previous experiences in producing the sound (Guenther, 2006). An efference copy of the motor command
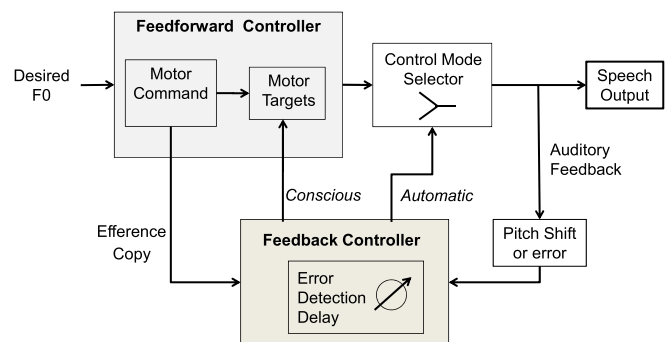

FIG. 7. (Color online) Vocal feedforward-feedback mode switching. Schematized model of system composed of feedforward and feedback controllers for voice F0 control. In this model, speaking begins in feedforward mode as default. The speaker then volitionally or automatically selects whether to continue operating in a feedforward mode or to switch to a feedback mode. In the feedforward mode, the presence of an external reference (e.g., piano) or something such as a learned stress pattern in speech may be used to guide the process of setting motor targets. When operating in the negative feedback mode, auditory feedback is used to compare and correct for errors between the desired and actual voice F0 produced.

J. Acoust. Soc. Am. 146 (6), December 2019

Patel et al. 4251

is sent from the feedforward controller to the feedback controller. Here, a comparison is made between the efference copy and the auditory feedback, a process that takes around 200 ms. The results of this comparison are sent to the control mode selector, and if an adjustment is required, the selector chooses the feedback mode, results as the subject attempts to match or follow a referent (vocal note or syllable). The results of both the feedforward and feedback controllers are sent to a selector, which allows for (volitional and automatic) selection of feedforward or feedback modes. Thus, following the delay, the volitional opposing control mechanism may be activated allowing a response that opposes the stimulus direction. This model differs from established models such as the Directions Into Velocities of Articulators model (Guenther, 2006) in that (1) there are two areas where feedback has input and (2) a selector exists to operationalize the choice of being in a feedforward mode or integrating feedback information.

Results of the present study also allow for a re-interpretation of results of previous studies. As previously noted (Patel *et al.*, 2014), the volitional following responses we report may be the same as the occasional following responses that others have reported in studies of the pitch-shift reflex. In that paradigm (e.g., Bauer *et al.*, 2006; Behroozmand *et al.*, 2012; Burnett *et al.*, 1998, 2008; Hain *et al.*, 2000), subjects were not instructed to volitionally change their voice $F0$ but to keep it steady and ignore changes in auditory feedback. It is possible that in the course of testing, a subject may occasionally pay greater attention to the feedback stimulus and inadvertently treat it as the referent and follow the direction of the stimulus (Fig. 6; Hain *et al.*, 2000). In an examination of the number of opposing and following responses that contribute to an averaged opposing response, Behroozmand and colleagues (2012) reported a nearly 50–50 split in the number of opposing and following responses in a set of 70–80 trials. Therefore, in studies of the pitch-shift reflex, the mix of opposing and following responses may reflect variable cognitive factors (e.g., attention) during the testing.

There are a few limitations of this study to keep in mind, such as the pitch-shifting process itself. Pitch determination is not without flaws. Even though the Eventide Eclipse Harmonizer attempts to shift pitch and the corresponding harmonic structure, this transformation may change the spectrum in ways that were not intended as well. Nevertheless, the Harmonizer has been used extensively in the music industry and for conducting pitch-shift research due to the quality of signal processing and near real-time feedback processing (delays informally observed less than 15 ms). Another limitation is that the variances observed between groups were not homogeneous, and although corrections were applied, heterogeneity of variances can increase the type I error. It is possible that this variability might have been reduced by instructing participants to match a particular pitch; however, it is not clear whether individuals without singing training would have been able to do such a task. Nevertheless, results show group differences in response magnitude and the presence of an early response in the opposing condition for all three groups. Our insights into the control of the voice are informed by different uses of the voice between all three groups of subjects.

V. CONCLUSIONS

In summary, it was found that people with a musical and tonal language background differ in the magnitude of their volitional opposing responses and the consistency of the held pitch compared to non-trained English speakers. Specifically, subjects of a tonal language (Mandarin) produced larger responses to pitch shifts, which may relate to their linguistic use of pitch. Singers, on the other hand, produced the smallest magnitudes of changes and held their pitch at a steadier level. These observations support the idea that greater musical training or the lexical use of tone enhances mechanisms of voice $F0$ control. These results may also reflect differences based on pitch goals of the two groups— singers learn to control pitch variability by minimizing deviations from an intended pitch, whereas Mandarin speakers may maximize variations in pitch in order to differentiate among tones in Mandarin, which would be linguistically advantageous. In addition, it was found that volitional following responses were produced with the same latency for the Mandarin speakers, trained vocalists, and people who are not musically trained and do not have a tonal language background. We propose that the neural mechanisms controlling the volitional following responses must be rather fixed and not easily amenable to change. Most importantly, it is evident that the control mechanisms for the following responses appear to be more readily accessed than the volitional opposing responses because most of the subjects produced an initial small following response that occurred before and perhaps delayed the onset of the opposing response.

Bailly, G. (**2003**). "Close shadowing natural versus synthetic speech," Int. J. Speech Technol. **6**, 11–19.

Bakker, K., and Brutten, G. J. (**1989**). "A comparative investigation of the laryngeal premotor, adjustment, and reaction times of stutterers and non-stutterers," J. Speech Hear. Res. **32**(2), 239–244.

Bauer, J. J., Mittal, 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.

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 Cognit. **84**(1), 97–108.

Behroozmand, R., Karvelis, L., Liu, H., and Larson, C. (**2009**). "Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation," Clin. Neurophysiol. **120**, 1303–1312.

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**(4), 2468–2477.

Belin, P. (**2006**). "Voice processing in human and non-human primates," Philos. Trans. Royal Soc. London Ser. B **361**(1476), 2091–2107.

Belin, P., and Zatorre, R. J. (**2003**). "Adaptation to speaker's voice in right anterior temporal lobe," Neuroreport **14**(16), 2105–2109.

Blakemore, S. J., Frith, C. D., and Wolpert, D. M. (**1999**). "Spatio-temporal prediction modulates the perception of self-produced stimuli," J. Cogn. Neurosci. **11**(5), 551–559.

Blakemore, S. J., Rees, G., and Frith, C. D. (**1998**). "How do we predict the consequences of our actions? A functional imaging study," Neuropsychologia **36**(6), 521–529.

Boersma, P., and Weenink, D. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345.

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**(6), 3153–3161.

Burnett, T. A., McCurdy, K. E., and Bright, J. C. (**2008**). "Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error," Exp. Brain Res. **191**(3), 341–351.

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (**2007**). "Mismatch negativity to pitch contours is influenced by language experience," Brain Res. **1128**(1), 148–156.

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**(2), 1157–1163.

Chen, Z., Liu, P., Wang, E. Q., Larson, C. R., Huang, D., and Liu, H. (**2012**). "ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization," Brain Lang. **121**(1), 25–34.

Cross, D. E., and Luper, H. L. (**1983**). "Relation between finger reaction time and voice reaction time in stuttering and nonstuttering children and adults," J. Speech Hear. Res. **26**(3), 356–361.

Dalziell, A. H., Welbergen, J. A., Igic, B., and Magrath, R. D. (**2015**). "Avian vocal mimicry: A unified conceptual framework," Biol. Rev. **90**(2), 643–668.

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**(1), 357–366.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., and Lowe, M. (**2004**). "Hemispheric roles in the perception of speech prosody," NeuroImage **23**(1), 344–357.

Goodale, E., and Kotagama, S. W. (**2006**). "Context-dependent vocal mimicry in a passerine bird," Proc. Biol. Sci. **273**(1588), 875–880.

Guenther, F. H. (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disorders **39**(5), 350–365.

Hafke, H. Z. (**2008**). "Nonconscious control of fundamental voice frequency," J. Acoust. Soc. Am. **123**(1), 273–278.

Hain, T. C., Burnett, T. A., Kiran, S., 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.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," Acoust. Res. Lett. Online **6**(3), 131–137.

Horii, Y. (**1979**). "Fundamental frequency perturbation observed in sustained phonation," J. Speech Hear. Res. **22**, 5–19.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech, Lang., Hear. Res. **45**(2), 295–310.

Houde, J. F., and Nagarajan, S. S. (**2011**). "Speech production as state feedback control," Front. Human Neurosci. **5**, 407–422.

Izdebski, K. (**1980**). "Effects of prestimulus interval on phonation initiation reaction times," J. Speech Hear. Res. **23**, 485–489.

Izdebski, K., and Shipp, T. (**1978**). "Minimal reaction times for phonatory initiation," J. Speech Hear. Res. **21**, 638–651.

Jones, J. A., and Keough, D. (**2008**). "Auditory-motor mapping for pitch control in singers and nonsingers," Exp. Brain Res. **190**(3), 279–287.

Kelley, L. A., and Healy, S. D. (**2010**). "Vocal mimicry in male bowerbirds: Who learns from whom?," Biol. Lett. **6**(5), 626–629.

Kelley, L. A., and Healy, S. D. (**2011**). "Vocal mimicry," Curr. Biol. **21**(1), R9–R10.

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**(2), 837–846.

Korzyukov, O., Tapaskar, N., Pflieger, M. E., Behroozmand, R., Lodhavia, A., Patel, S., Robin, D. A., and Larson, C. (**2015**). "Event related potentials study of aberrations in voice control mechanisms in adults with

attention deficit hyperactivity disorder," Clin. Neurophysiol. **126**(6), 1159–1170.

Kuhl, P. K., and Meltzoff, A. N. (**1996**). "Infant vocalizations in response to speech: Vocal imitation and developmental change," J. Acoust. Soc. Am. **100**(4 Pt 1), 2425–2438.

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007**). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech, Lang., Hear. Res. **50**(4), 913–927.

Leonard, R. J., Ringel, R., Horii, Y., and Daniloff, R. (**1988**). "Vocal shadowing in singers and nonsingers," J. Speech Hear. Res. **31**(1), 54–61.

Leonard, R. J., and Ringel, R. L. (**1979**). "Vocal shadowing under conditions of normal and altered laryngeal sensation," J. Speech Hear. Res. **22**, 794–817.

Lilly, J. C. (**1965**). "Vocal mimicry in tursiops: Ability to match numbers and durations of human vocal bursts," Science **147**(3655), 300–301.

Liu, H., Auger, J., and Larson, C. R. (**2010a**). "Voice fundamental frequency modulates vocal response to pitch perturbations during English speech," J. Acoust. Soc. Am. **127**(1), EL1–EL5.

Liu, H., Russo, N. M., and Larson, C. R. (**2010b**). "Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study," J. Acoust. Soc. Am. **127**(2), 1042–1046.

Nagao, K., McCurdy, K. E., and Burnett, T. A. (**2008**). "Age effects of pitch-shifted auditory feedback on reflexive and volitional voice F0 control," J. Acoust. Soc. Am. **123**(5), 3072–3072.

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.

Ning, L. H., Loucks, T. M., and Shih, C. (**2015**). "The effects of language learning and vocal training on sensorimotor control of lexical tone," J. Phon. **51**, 50–69.

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**, 46–46.

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., and Robin, D. A. (**2012**). "Understanding the neural mechanisms involved in sensory control of voice production," NeuroImage **61**(1), 314–322.

Patel, S., Lodhavia, A., Frankford, S., Korzyukov, O., and Larson, C. R. (**2016**). "Vocal and neural responses to unexpected changes in voice pitch auditory feedback during register transitions," J. Voice **30**(6), 772.e33–772.e40.

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.

Peschke, C., Ziegler, W., Kappes, J., and Baumgaertner, A. (**2009**). "Auditory-motor integration during fast repetition: The neuronal correlates of shadowing," NeuroImage **47**, 392–402.

Ramanarayanan, V., Parrell, B., Goldstein, L., Nagarajan, S., and Houde, J. F. (**2016**). "A new model of speech motor control based on task dynamics and state feedback," *Paper presented at the Interspeech 2016*, San Francisco, CA.

Reiss, D., and McCowan, B. (**1993**). "Spontaneous vocal mimicry and production by bottlenose dolphins (Tursiops truncatus): Evidence for vocal learning," J. Comp. Psychol. **107**(3), 301–312.

Richards, D. G., Wolz, J. P., and Herman, L. M. (**1984**). "Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncates," J. Comp. Psychol. **98**(1), 10–28.

Shipp, T., and Izdebski, K. (**1975**). "Vocal frequency and vertical larynx positioning by singers and nonsingers," J. Acoust. Soc. Am. **58**(5), 1104–1106.

Shockley, K., Sabadini, L., and Fowler, C. A. (**2004**). "Imitation in shadowing words," Percept. Psychophys. **66**(3), 422–429.

Takeuchi, A. H., and Hulse, S. H. (**1993**). "Absolute pitch," Psychol. Bull. **113**(2), 345–361.

J. Acoust. Soc. Am. **146** (6), December 2019

Patel *et al.* 4253

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," NeuroImage **39**, 1429–1443.

Watson, B. C. (**1994**). "Foreperiod duration, range, and ordering effects on acoustic LRT in normal speakers," J. Voice **8**(3), 248–254.

Wilbrecht, L., and Nottebohm, F. (**2003**). "Vocal learning in birds and humans," Ment. Retard Dev. Dis. Res. Rev. **9**(3), 135–148.

Wolpert, D. M. (**1997**). "Computational approaches to motor control," Trends Cognit. Sci. **1**, 209–216.

Wolpert, D. M., and Miall, R. C. (**1996**). "Forward models for physiological motor control," Neural Networks **9**(8), 1265–1279.

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**(2), 1168–1178.

Zatorre, R. J. (**2001**). "Neural specializations for tonal processing," Ann. N. Y. Acad. Sci. **930**, 193–210.

Zatorre, R. J. (**2003**). "Music and the brain," Ann. N. Y. Acad. Sci. **999**, 4–14.

Zatorre, R. J., and Gandour, J. T. (**2008**). "Neural specializations for speech and pitch: Moving beyond the dichotomies," Philos. Trans. Royal Soc. London Ser. B **363**(1493), 1087–1104.

4254   J. Acoust. Soc. Am. **146** (6), December 2019

Patel *et al.*