Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback

Sona Patel,^{a)} Cristina Nishimura, Anjali Lodhavia, and Oleg Korzyukov
Northwestern University, Evanston, Illinois 60208

Amy Parkinson and Donald A. Robin
University of Texas Health Sciences Center, San Antonio, Texas 78229

Charles R. Larson
Northwestern University, Evanston, Illinois 60208

(Received 1 July 2013; revised 11 January 2014; accepted 25 March 2014)

Previous research has shown that vocal errors can be simulated using a pitch perturbation technique. Two types of responses are observed when subjects are asked to ignore changes in pitch during a steady vowel production, a compensatory response countering the direction of the perceived change in pitch and a following response in the same direction as the pitch perturbation. The present study investigated the nature of these responses by asking subjects to *volitionally* change their voice fundamental frequency either in the opposite direction (“opposing” group) or the same direction (“following” group) as the pitch shifts (± 100 cents, 1000 ms) presented during the speaker’s production of an /a/ vowel. Results showed that voluntary responses that followed the stimulus directions had significantly shorter latencies (150 ms) than opposing responses (360 ms). In addition, prior to the slower voluntary opposing responses, there were short latency involuntary responses that followed the stimulus direction. These following responses may involve mechanisms of imitation or vocal shadowing of acoustical stimuli when subjects are predisposed to respond to a change in frequency of a sound. The slower opposing responses may represent a control strategy that requires monitoring and correcting for errors between the feedback signal and the intended vocal goal.
© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4870490>]

PACS number(s): 43.71.Rt, 43.70.Mn, 43.66.Hg, 43.64.Bt [MAH]

Pages: 3036–3044

I. INTRODUCTION

Motor control of the voice is heavily dependent on auditory feedback. Once vocalization begins, if an error is detected between the ongoing vocalization and voice auditory feedback, the disparity in the acoustical properties of the produced vocalization and the auditory feedback trigger a compensatory vocal response. This compensatory response has been demonstrated in previous studies in which the voice pitch is temporarily altered during the production of vowels (Burnett *et al.*, 1998; Donath *et al.*, 2002; Natke *et al.*, 2003; Hafke, 2008; Behroozmand *et al.*, 2009; Hawco and Jones, 2009). In addition, compensatory responses have also been observed in studies that shift the first or second formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006; Cai *et al.*, 2010). Typically, compensation to pitch shifts is exhibited as a corrective “opposing” response to counter the direction of the shift, indicating a closed-loop negative feedback system (Hain *et al.*, 2000; Natke *et al.*, 2003). Because these responses happen without intent by the speaker and people are unable to suppress them (Zarate and Zatorre, 2008), they are thought to be involuntary in nature. However, a fairly large proportion of responses is imitating or “following” responses that match the direction of the

shift. From several studies that have reported the numbers of following responses to pitch-shifted voice auditory feedback, they constituted between 1.5% and 39% of all responses (Burnett *et al.*, 1998; Larson *et al.*, 2000; Hain *et al.*, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005; Chen *et al.*, 2007; Larson *et al.*, 2007; Liu and Larson, 2007; Larson *et al.*, 2008; Liu *et al.*, 2011). These variable percentages were due to different experimental conditions such as variations in the shift magnitudes. The mechanisms underlying the division in response qualities are not yet understood.

Most studies examining pitch-shifting in vowels have examined the involuntary responses obtained when speakers are instructed to ignore changes in pitch. Only a couple of studies have examined voluntary responses to pitch-shifted voice feedback (Larson, 1998; Hain *et al.*, 2000). Both studies have commented on the complex nature of the response patterns obtained that consisted of multiple responses to a single pitch-shift stimulus varying in onset time or “latency.” In the cases where at least two responses were observed, the first response (involuntary) occurred around 100–150 ms, and the second response (voluntary) occurred between 250 and 600 ms. The earlier responses were considered involuntary because they were produced without the overt attempt to make them, and they generally opposed the direction of the stimulus. Nevertheless there were also occurrences where the early latency responses followed the stimulus direction and were of a large magnitude, suggesting they could be voluntarily produced. The authors also commented

^{a)} Author to whom correspondence should be addressed. Present address: Seton Hall University, South Orange, NJ 07079. Electronic mail: Sona.Patel@shu.edu

on the observation that the instruction on how to respond to pitch shift stimuli had a significant effect on response latency and magnitude. The observation that the instructions affected response characteristics is not unusual considering that studies of voluntary voice reaction times (Izdebski and Shipp, 1978; Cross and Luper, 1983; Shipp *et al.*, 1984; Bakker and Brutten, 1989; Watson, 1994) and vocal shadowing studies (Leonard and Ringel, 1979; Horii, 1984; Leonard *et al.*, 1988; Bailly, 2003; Shockley *et al.*, 2004; Peschke *et al.*, 2009) lead to responses with latencies that are similar to the responses reported by Hain *et al.* (2000).

The present study grew out of a larger one that was intended to investigate the effects of training people to voluntarily respond to pitch-shifted voice feedback. In that study, we implemented a similar task to that in Hain *et al.* (2000), but participants were divided into two groups: One group of participants (the “opposing” group) was instructed to change their F0 in the opposite direction of the pitch-shift the pitch-shift stimulus or “oppose the pitch shift” and the other (the “following group”) was instructed to change their F0 in the same direction of the stimulus or “follow the pitch shift.” Participants in both groups vocalized and then produced voluntary changes in their voice F0 to either oppose or follow the direction of a pitch shift stimulus. We measured the latency and magnitudes of the voluntary responses and possible involuntary responses.

II. METHODS

A. Participants

Participants were recruited from Northwestern University. Individuals who met the following criteria were eligible to participate in this study: right-hand dominant, native speaker of American English, normal hearing [at octave intervals from 250 to 8000 Hz at 20 dB hearing level (HL); ANSI S3.21-2004 (ANSI, 2004)], a score of at least 90% (18 of 20) correct on two tests of central auditory processing or CAP including the Duration Pattern Sequence test (Musiek *et al.*, 1990) and the Pitch Pattern Sequence test (Musiek and Pinheiro, 1987), and report having no history of neurological, speech, or language disorders. The hearing screening and CAP tests were performed in the laboratory prior to testing. Also participants were required to have minimal vocal training defined as less than three years of vocal training and report not currently singing in a group on a regular basis (three times a week or more). Of the 22 participants, 11 participants were randomly assigned to the “opposing” group (3 male, 8 female; mean: 20.0 yr, standard deviation or SD: 1.2 yr) and 11 participants (3 male, 8 female; mean: 21.3 yr, SD: 1.6 yr) were assigned to the “following” group.

B. Procedures

The present experiment consisted of two parts, *baseline* testing followed by testing of *volitional* responses to the stimuli. Participants were seated in a double-walled, sound-treated booth. A visual display was presented on a computer screen that instructed the participant to either “get ready” or “say aah” at which time they were asked to vocalize a

sustained /a/ sound. A progress bar on a computer monitor located in front of them indicated the length of time for each vocalization. During each vocalization, perturbations in the auditory feedback were presented by shifting the participant’s pitch upward or downward (randomized direction) by 100 cents (100 cents = 1 semitone). An approximately equal number of upward and downward pitch shifts was collected.

For the baseline test, participants vocalized for 5.5 s and were told to ignore any changes in the sound of their voice through the headphones. During each vocalization, five 200-ms shifts were presented. The first shift in each vocalization was presented randomly between 700 and 1000 ms after voice onset with the subsequent shifts occurring at randomized times between 500 and 700 ms following the previous shift offset. There was a 2 s pause between each vocalization. A total of 260 trials were recorded in 52 vocalizations (52 vocalizations $\times$ 5 pitch shifts per vocalization). Between each block of trials, participants rested and drank water as needed.

For the volitional test, participants vocalized for 2 s and were presented with a single 1000-ms long pitch-shift per vocalization (either 100 cents down or up). Participants were instructed to actively change their voice F0 to either oppose (opposing group) or follow (following group) the direction of the actual shift and maintain the modified level for the remainder of the vocalization. For example, when participants in the opposing group were given a downward shift, they would modify their F0 to be higher than their normal F0, whereas participants in the following group would reduce their F0 to be lower than normal. Participants were free to change the F0 of their voice by whatever magnitude they desired. Because we were interested in the accuracy of response direction rather than speed of the change in voice F0, the participants were not instructed on how fast they should respond to the stimuli, only that the responses should be in the direction specified by the instructions. Participants were first instructed how to respond to the stimuli, and then they performed a short practice session of 10 trials where they changed the F0 of their voice when they heard a pitch-shift stimulus. The pitch-shift occurred randomly between 500 and 1000 ms after voice onset. A total of 208 trials were collected in four blocks of 52 vocalizations for each participant. Participants were given a short break after each block. The experiment lasted 1 h. Participants were given water as often as needed and compensated for their participation.

C. Instrumentation

All vocalizations were recorded using an AKG boomset microphone (C420; AKG, location). During each vocalization, the participant’s pitch was shifted using an Eclipse Harmonizer (Eventide, Little Ferry, NJ), amplified with a 10 dB gain using a Mackie mixer (Model No. 1202), and presented as real-time feedback using a Sennheiser headset during the main test. The vocalizations, modified voice feedback signal, and control pulses (used as an indicator of the direction of the pitch-shift) were digitized at 10 kHz, low-passed filtered at 5 kHz, and recorded using LABCHART software (AD Instruments, Colorado Springs, CO). MIDI software (Max/MSP v. 5.0) was used to present the visual

display, to present the vocalizations as feedback and to control characteristics of the pitch-shift including direction randomization, timing, and magnitude.

D. Data analysis

The data were analyzed using scripts designed in IGOR PRO (Wavemetrics, Inc., Lake Oswego, OR). This software made use of the F0 detection in PRAAT (Boersma and Weenink, 2001). First, the vocalizations were aligned with a transistor-transistor logic (TTL) pulse that signified the onset of the pitch-shift stimulus. Next, the voice data were segmented into trials and converted into F0 contours. For analysis of the baseline test, each trial was segmented into an 800 ms voice segment consisting of 200 ms pre-shift and a 600 ms post-shift segment. Then a number of preprocessing steps were performed to minimize outliers in each trial, including normalization by setting the mean baseline voice pitch to 0 cents, removal of extreme values in the vocalization wave prior to the pitch-shift (for $threshold = 20$ cents, where $\max \text{ cents} > threshold$ and $\min \text{ cents} < -threshold$ were rejected), and removal of extreme values in the entire vocalization and feedback waves (for $threshold = 1000$ cents, where the whole wave was rejected if $\max \text{ cents} > threshold$ or $\min \text{ cents} < -threshold$). After the preprocessing steps were taken, the F0 responses were averaged by first sorting the responses according to the stimulus and response direction to eliminate the contamination of averaging responses by including both upward and downward responses into the average (Behroozmand *et al.*, 2012). Thus the averaged responses were composed only of responses that either increased or decreased following the stimulus.

Once the outlier removal process was completed, the vocalization and feedback waves were averaged for each condition (+100-cent shifts, -100-cent shifts) within a participant. From these averaged segments, the time at which the response began (the latency) was measured from the pitch shift onset (specifically, the point at which the pitch of the voice response first exceeded two standard deviations of the mean amplitude in the window of -200 to 0 ms, where 0 indicates pitch shift onset). The response magnitude was

measured as the time at which the response was maximal. To identify differences in the latency and magnitude due to group and stimulus direction, two-way repeated measures analysis of variances (ANOVA) were performed in SPSS (v.17; SPSS Inc., Chicago, IL). Finally, the vocalization and feedback waves were grand averaged across participants by group.

Analysis of the volitional responses was done similarly with the exception that outliers representing an incorrect response (e.g., a following response instead of an opposing response) were removed before averaging results for each subject. The dependent variables measured were the response latency and magnitude. For this, the analysis window was extended to 1400 ms (1200 ms post-stimulus) to measure the response magnitudes and to determine if subjects reached a plateau in their altered voice F0 prior before the trial ended. For measuring the response latencies, we used a criterion that responses must exceed a value of two SDs of the pre-stimulus mean.

III. RESULTS

A. Baseline results

Figure 1 illustrates grand averaged responses for the opposing and following groups according to stimulus condition. The graphs show the similarity in responses for both subject groups by combining opposing and following response directions in each graph. Table I shows the mean onset times for both groups in response to upward and downward pitch shift stimuli. The differences in the mean onset time (118 ms, SD: 69) for the following and the opposing groups (130 ms, SD: 85) were not significantly different [ANOVA, $F(1,36) = 0.173$, $p < 0.7$]. The mean response magnitudes for the following group, (14.5 cents, SD: 8) and for the opposing group (14.6 cents, SD: 7) were also not significantly different [ANOVA, $F(1,42) = 0.0$, $p < 0.9$]. As is evident from Fig. 1 and our measures obtained from individual averages across subjects, the average response magnitudes in the baseline condition did not differ between the two subject groups or between the response directions.

FIG. 1. Traces representing group averages of voice F0 responses to pitch-shifted auditory feedback in the baseline condition. Responses in the downward direction (to upward shifts for the opposing group and to downward shifts in the following group) are shown on the left. Responses in the upward direction (to downward shifts for the opposing group and upward shifts in the following group) are shown on the right. Gray traces are from the following group and black traces from the opposing group. Brackets at the bottom indicate time and direction of the pitch shift stimulus. Error bars represent standard deviation of the mean.

TABLE I. Mean onset times in milliseconds (and standard deviations or SD) of the responses made by both groups (opposing group, following group) for the baseline and volitional test conditions for both upward and downward stimuli. Standard deviations in parentheses.

	Opposing group		Following group	
	Up	Down	Up	Down
Baseline	114 (71)	144 (102)	133 (71)	104 (68)
Volitional	370 (154)	303 (132)	171 (96)	187 (110)

B. Voluntary response onset

Prior to data analysis, 9% of responses for the following group and 10% of the opposing group responses were removed because they were made in the wrong direction. The grand averages of the (within-subject averaged) voice F0 contours across participants for each group are shown in Fig. 2. Figure 2(A) shows the grand average of responses for the opposing and following groups for an upward pitch-shift stimulus. For this condition, the following group produced an upward response (latency: 171 ms), and the opposing group made downward responses (latency: 370 ms). Figure 2(B) shows average responses to downward pitch-shift stimuli. For this condition the following group made a downward response (latency: 187 ms), and the opposing group made an upward response (latency: 303 ms, see Table I).

A multivariate ANOVA was conducted to examine the effect of two fixed factors on response onset time and magnitude, including group (opposing, following) and stimulus direction (up, down). Our dependent variables were normally distributed according to a Kolmogorov–Smirnov test (K-S $Z=0.530$, $p=0.941$ for magnitude and K-S $Z=0.678$, $p=0.747$ for onset time). There was homogeneity of variance between groups as assessed by Levene’s test for equality of error variances [$F(3,35)=1.152$, $p=0.342$ for magnitude and $F(3,35)=0.434$, $p=0.730$ for onset time]. Results show a significant main effect of group on response onset time, $F(1, 35)=14.862$, $p<0.001$, but not stimulus direction, $F(1, 35)=0.397$, $p=0.533$. The interaction of group and stimulus direction was also not significant, $F(1,35)=1.014$, $p=0.321$. No significant differences in the responses magnitude were observed between groups [$F(1,35)=0.207$, $p=0.652$] or between the two stimulus directions [$F(1,35)=0.027$, $p=0.870$]. However, there was a significant interaction between group and direction [$F(1,35)=6.873$, $p<0.05$]. The following group showed significantly larger response magnitudes for upward responses (382 cents, SD: 165) compared to (absolute values of) downward response magnitudes (253 cents, SD: 144), paired samples t -test $t(10)=-2.791$, $p<0.05$. The peak response magnitudes for the opposing group were marginally significant [paired samples t -test $t(10)=2.190$, $p=0.053$] between upward responses (i.e., responses to downward shifts; 353 cents, SD: 200) and (absolute values of) downward responses (i.e., responses to upward shifts; 219 cents, SD: 116).

FIG. 2. Group-averaged voice F0 contours in the voluntary response condition. Black traces represent the opposing group and gray traces the following group. (A) illustrates averages for upward stimuli and (B) shows averages for downward stimuli. Brackets at the bottom of each panel indicate stimulus timing and direction. (C) and (D) show the same data as in (A) and (B) with expanded scales to show early involuntary responses (see arrows) prior to the voluntary responses for the opposing group. Error bars represent standard deviation of the mean.

C. Complex response onsets

We observed that some participants in the opposing group (eight for the upward stimuli and nine for the downward stimuli; a subset of seven produced both) produced a small response that followed the stimulus direction and occurred before the intended opposing response, potentially causing the delay in the main response. The mean responses for the opposing and following groups are shown with expanded scales in Figs. 2(C) and 2(D). The early and involuntary following responses [early upward response in Fig. 2(C) and early downward response in Fig. 2(D)] made by the opposing group (black lines) are indicated with arrows. These early involuntary responses occurred for only the opposing group prior to the intended opposing responses [volitional downward response in Fig. 2(C) and volitional upward response in Fig. 2(D)]. The mean onset time for the early following response in the downward direction [Fig. 2(D)] was 120 ms and in the upward direction [Fig. 2(C)] was 140 ms. The peak magnitude of the early upward response [Fig. 2(C)] was 12 cents and of the early downward response [Fig. 2(D)] was -4 cents.

The onset times of these early following responses made by the opposing group are nearly identical to the volitional responses made by the following group. The mean onset times for both the opposing and following groups during the baseline and volitional tests are reported in Table I.

IV. DISCUSSION

The present study provides important new information regarding the control of the voice based on auditory feedback. The results show new features of vocal responses to pitch-shifted voice feedback. Most importantly we showed that when participants are instructed to *voluntarily* match the direction of a pitch perturbation in their voice feedback (following group), they respond about 200 ms faster than participants who are asked to change their F0 in the opposite direction of the shift (opposing group). A second major finding that may be linked to the first is that most participants in the opposing group produced a small response that followed the direction of the pitch shift just before the intended opposing response. We suggest these small initial responses are involuntary because they were not made with overt intent. Instead these responses occurred when attempting to make a response in the opposite direction. The opposing group also showed longer latencies for the voluntary response, suggesting that the neural circuitry involved in producing an overt opposing action requires more processing time and therefore may be more complex than the mechanisms controlling the voluntary following responses. It is unlikely that these differences are due to inherent differences in the two subject groups because both groups produced identical response latencies (of the pitch-shift reflex) during the baseline testing condition.

These results suggest the presence of a new type of voice reflex that to our knowledge has not been previously reported. Previous studies have shown that when subjects sustain a constant voice F0 with no intention to change it, pitch-shifted voice auditory feedback elicits a compensatory

response (Burnett *et al.*, 1998; Larson *et al.*, 2000; Hain *et al.*, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005; Chen *et al.*, 2007; Larson *et al.*, 2007; Liu and Larson, 2007; Larson *et al.*, 2008; Liu *et al.*, 2011). It was suggested that this response was a reflex because it was produced without intention, was difficult to prevent (Zarate and Zatorre, 2008), and usually was in the opposite direction to the stimulus. It was further suggested that this reflex functioned as part of a negative feedback control system that was optimal for stabilizing voice F0 at a desired level (Hain *et al.*, 2000). Results of the present study suggest that if the *intention* of the subject is to volitionally respond to a change in voice pitch feedback by changing their voice F0, an involuntary change in F0 is elicited that changes F0 in the *same direction* as the pitch-shifted feedback.

It is also important to note that the involuntary following responses were only apparent in the opposing subject group (Fig. 2). Moreover, the fact that the involuntary following responses had the same latencies as the voluntary following responses suggests that the failure to observe the involuntary responses in the following group is because they were integrated with the voluntary following responses. This suggestion further implies that one of the reasons for the short latency voluntary following responses is because the involuntary following response was initiated and quickly became part of the voluntary following response. In other words, if a person is predisposed to change his/her voice F0 to match that of a pitch-shifted version of his/her own voice, there is an immediate and automatic response to change voice F0 to match the change in pitch. This mechanism assists the subject in making a very rapid change in voice F0. This same mechanism may underlie the propensity for singers to match musical accompaniment such as from a piano.

This explanation of the voluntary following responses may also explain the relatively slow opposing responses. Namely, as seen in Fig. 2, the opposing responses did not begin until after the involuntary following response occurred. In other words, the initial involuntary following response delayed the onset of the intended opposing response. However, this may not be the only factor that explains the longer latency of the opposing responses. It should be noted that of the five participants in the opposing group that did not produce an involuntary following response, their responses had the same latency as those in the opposing group who did produce a involuntary following response. That is, even without first producing a involuntary following response, the opposing responses still had a longer latency (onset time) than the voluntary following responses.

An unexpected finding related to the volitional response magnitudes was that the downward responses were significantly smaller than the upward responses. Because no instructions were given to the subjects on how much they should change their voice F0 in response to the stimuli, we assume that all subjects adopted a somewhat similar strategy and avoided changes in F0 that required great effort. The downward responses may have been smaller than the upward responses because subjects generally started their vocalization at or near their normal conversational level. At this

level, most people can elevate their voice F0 by a much greater amount than they can lower it. Also, most laryngeal muscles work to increase F0 by contraction and decrease F0 by relaxation (Hirano and Ohala, 1969; Hirose and Gay, 1972; Ludlow and Lou, 1996). Beyond the relaxed level these muscles are incapable of lowering F0 further. To lower one's voice below conversational level typically involves contraction of the strap muscles (Roubeau *et al.*, 1997). Even with these muscles greater effort may be needed to lower the voice much below a few semitones.

In previous studies using the pitch-shift paradigm, it was suggested that a negative feedback system controlled the responses and functioned to correct for errors in F0 production, thereby stabilizing voice F0 at a desired level (Burnett *et al.*, 1998). Nevertheless, several studies have reported that some responses followed the stimulus direction, which in effect would cause de-stabilization of F0 control (Burnett *et al.*, 1998; Larson *et al.*, 2000; Hain *et al.*, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005; Chen *et al.*, 2007; Larson *et al.*, 2007; Liu and Larson, 2007; Larson *et al.*, 2008; Liu *et al.*, 2011). Recently, Behroozmand *et al.* (2012) suggested that one mechanism initiating following responses was that they could have resulted from vocal tremor prior to the pitch shift that caused the F0 contour to continue in a following direction after stimulus onset. Another suggestion was that following responses could possibly represent a form of vocal mimicry. While neither of these suggestions can provide a definitive explanation for following responses, results of the present study provide additional information that may explain them. To this end, it is important to note that the following responses occurred with the instruction to voluntarily change voice F0 while subjects were listening to vocal sounds. Thus it seems that the crucial factors leading to the voluntary and involuntary following responses in the present study were the intention to voluntarily change voice F0 and the presence of an acoustical cue, in this case, the feedback of the participant's voice. It is also important to note that the latency of the voluntary following responses was about 200 ms shorter than those that were made in the opposite direction of voice pitch feedback changes. Therefore it is suggested that the volitional following responses reported here represent vocal imitation, shadowing, or mimicry, while the opposing responses represent more complex responses, perhaps involving negative feedback circuitry that corrects for errors in production.

A follow-up question related to this proposal is whether the involuntary following responses reported in studies of the pitch-shift reflex in which participants are sustaining a steady phonation with no intention to voluntarily change F0 in response to a pitch-shift stimulus (e.g., Burnett *et al.*, 1998) reflect the same mechanism as the voluntary following responses of the present study. Unfortunately, we cannot provide a definitive answer to this question. However, it may be safe to suggest that in the process of producing several short trials of vocalization during a testing session, some individuals may subtly change their vocal control mechanisms and occasionally allow their vocal control system to drift into an imitative mode rather than maintaining a feedback-based motor control mode. This question could

possibly be answered by the design of experiments in which there are more stringent attempts to control participants' vocal control mechanisms during testing.

One interpretation of the short response latencies observed in the voluntary following condition in the present study is that they may reflect the same mechanisms as those previously reported for vocal or verbal shadowing (Leonard and Ringel, 1979; Horii, 1984; Leonard *et al.*, 1988; Bailly, 2003; Shockley *et al.*, 2004; Peschke *et al.*, 2009). Vocal or verbal shadowing is the repetition of changes in sounds or speech segments spoken by another person, and they are made with a delay varying between 70 and 250 ms. The speed of vocal shadowing (Horii, 1984) is on the same order of magnitude as the voluntary following responses reported here. However, our data cannot be directly compared with vocal shadowing because individuals in our experiment made changes in the F0 of their voice in response to changes in their own voice auditory feedback not an artificial tone. Nevertheless, the response latencies reported here are of the same order of magnitude as those reported by Horii for vocal shadowing, and they are also in the same direction as the stimulus change and were made voluntarily. Leonard *et al.* (1988) and Jafari *et al.* (1989) compared the speed of singers and nonsingers to adjust the pitch of their voice to a target tone for both raising and lowering F0 starting from a number of different F0 levels. They found the speed of response onset in some cases was similar to the values we report. In a somewhat different approach, Nudelman *et al.* (1992) measured the speed of vocal F0 adjustment of people who do and do not stutter and again found similarities to the timing we report. Taken together, our data are very similar to the previously reported data on shadowing, suggesting that the speed of vocal adjustment in shadowing and the voluntary following responses we describe may be controlled by the same or similar mechanisms.

The measures of the present study can also be compared with studies of vocal reaction times. In the vocal reaction time experiments, subjects typically initiate vocalization as fast as possible following a stimulus such as an auditory cue. Results of these experiments reveal voice reaction times in the range of 150–250 ms (Izdebski and Shipp, 1978; Cross and Luper, 1983; Shipp *et al.*, 1984; Bakker and Brutten, 1989; Watson, 1994). These studies indicate that people are capable of initiating or volitionally changing their voice (F0 or amplitude) within about 200 ms of a stimulus. Although these reaction times are of the same order of magnitude as measures of the pitch-shift reflex, they may not necessarily involve the same neural mechanisms. Intentional behaviors including the voice are generally thought to involve prefrontal mechanisms (Koval *et al.*, 2011; Schafer and Moore, 2011). Reports using functional magnetic resonance imaging (fMRI) or electrocorticography (ECoG) methods suggest the superior temporal gyrus and premotor cortex are involved in the pitch-shift reflex (Parkinson *et al.*, 2012; Greenlee *et al.*, 2013). Thus far, there is no evidence that prefrontal cortex is involved in responses to pitch-shift stimuli. Thus the speed of responses does not mean that the mechanisms producing them are the same; just that the time is the same. Further studies are needed to determine the neural mechanisms

underlying the role of auditory feedback in voice control, both for automatic, i.e., involuntary behavior and volitional control of the voice.

Despite the varying methodology between the present report and previous studies on verbal or vocal shadowing, all seem to agree that there are unique neural mechanisms involved in changing speech or voice (F0) to match an auditory cue such as a change in the pitch of one's own voice (Leonard and Ringel, 1979; Horii, 1984; Leonard *et al.*, 1988; Bailly, 2003; Shockley *et al.*, 2004; Peschke *et al.*, 2009; Nudelman *et al.*, 1992). Indeed, because many animals including humans learn vocalization and speech by modeling their productions from their cohorts or parents, there may be a tendency to mimic or imitate vocalizations of others (Bailly, 2003; Peschke *et al.*, 2009). In birds and other animals, the tendency to mimic vocalizations of other birds is quite common (Richards *et al.*, 1984; Reiss and McCowan, 1993; Shockley *et al.*, 2004; Zollinger and Suthers, 2004; Knornschild *et al.*, 2010) and may reflect a biological mechanism as a step in the evolution of language in humans (Kuhl and Meltzoff, 1996; Wilbrecht and Nottebohm, 2003). Thus the rapid latency of the following responses in the present study may reflect a more primitive, but highly conserved tendency, to imitate the vocalizations of others.

From the perspective of neural modeling, the opposing responses seem to be controlled by a negative feedback system, whereas the following responses may be controlled by a volitional or mimicry system. Negative feedback systems are slow because of the time necessary to compare the intended output with sensory feedback and then make a corrective response based on the comparison. In general, responses generated by negative feedback control mechanisms are made in the opposite direction to the stimulus, i.e., the responses are compensatory to the stimuli. Thus both the timing and direction of the responses made by the volitional opposing group in the present study suggest the responses were made by a negative feedback control system.

On the other hand, the volitional following responses we report here seem to be driven by other mechanisms. It was previously suggested by Hain *et al.* (2000, their Fig. 5) that if a vocalist uses an internal referent such as memory to control voice F0, a perturbation in auditory feedback results in a response to cancel the perturbation and thereby correct for the error (opposing response). If the vocalist treats the auditory feedback signal as the referent, e.g., a piano note and if there is a change in the note, a following response is initiated that attempts to reduce the difference between one's own vocalization and the referent. Operation in both modes can be considered as a negative feedback control system; the only difference is whether the referent is internal, such as memory or external, such as a piano note. In one case, the vocalist is attempting to control the voice based on memory, such as a previously voiced note, or in the other case based on an external referent such as a piano. It may be speculated that neural mechanisms of mimicry or shadowing operate somewhat in the latter mode. That is, during development, children attempt to produce sounds like those heard in their environment or by their parents. Repeated attempts to produce the same speech or vocal sounds are made by correcting for productions that

do not match the referent. Through a long process of learning, the differences between the child's productions and that of their parents are gradually reduced to zero. This same method of learning has been conceptualized for speech learning in the DIVA model (Guenther, 1994, 1995; Guenther *et al.*, 2004; Guenther, 2006).

Finally, it is worth noting that the differences in volitional response latencies based on response direction are not unique to the vocal system. In the oculomotor system, it is well known that prosaccadic eye movements (in the same direction of stimulus similar to the following responses) are about 200 ms faster than antisaccades, which are in the opposite direction (similar to the opposing responses) (Seidlits *et al.*, 2003). Although the mechanisms of voice F0 direction change and eye movement control may be quite different, the timing similarities between these two systems support our contention that the latency differences in voice response direction we report represent important differences in their neural control mechanisms.

V. CONCLUSIONS

The present study has shown that when participants are instructed to voluntarily respond to perturbations in the pitch auditory feedback during vowel production, participants respond in the correct direction. However, responses that follow the stimulus direction have a much shorter latency (approximately 150 ms) than the responses that oppose the stimulus direction (approximately 350 ms). In addition, prior to the opposing direction responses, there were small responses that followed the stimulus direction. These responses appear to be involuntary and have the same latency as the volitional following responses. The similarity in latency of the involuntary and volitional following responses suggests that the two responses are integrated into a single response as subjects attempt to match a change in voice pitch auditory feedback. The difference in response latencies between the volitional following and opposing responses suggests they make use of different neural processes. The following responses may utilize a control mode wherein the pitch-shift stimulus is treated as a referent that the vocal system attempts to match. The opposing responses may be controlled by a negative feedback control system in which errors that do not match an internal referent are corrected. In the case of this experiment, by instructing the subjects to either oppose or follow the direction of the pitch-shift stimulus, we predisposed them to either treat the pitch-stimulus as the referent or their own internal referent of voice pitch. In both cases, subjects attempted to control their voice F0 by canceling the difference in their vocal F0 and the referent. It is also suggested that the mechanisms for voice control in which subjects attempt to match the pitch-shift stimulus are similar to vocal shadowing or mimicry.

ACKNOWLEDGMENTS

This research was supported by NIH Grant Nos. 1R01DC006243 and T32 DC009399. The authors would like to thank Chun Liang Chan for his help with the computer programming.

- ANSI (2004). S3.21, *Methods for Manual Pure-Tone Threshold Audiometry* (Acoustical Society of America, New York).
- Bailly, G. (2003). "Close shadowing natural versus synthetic speech," *Int. J. Speech Tech.* **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**, 239–244.
- 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**, 2468–2477.
- 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**, 3153–3161.
- Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (2010). "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization," *J. Acoust. Soc. Am.* **128**, 2033–2048.
- 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.
- 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**, 356–361.
- Donath, T. M., Natke, U., and Kalveram, K. T. (2002). "Effects of frequency-shifted auditory feedback on voice F0 contours in syllables," *J. Acoust. Soc. Am.* **111**, 357–366.
- Greenlee, J. D., Behroozmand, R., Larson, C. R., Jackson, A. W., Chen, F., Hansen, D. R., Oya, H., Kawasaki, H., and Howard, M. A., III (2013). "Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex," *PLoS One* **8**, e60783.
- Guenther, F. H. (1994). "A neural network model of speech acquisition and motor equivalent speech production," *Biol. Cybern.* **72**, 43–53.
- Guenther, F. H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," *Psychol. Rev.* **102**, 594–621.
- Guenther, F. H. (2006). "Cortical interactions underlying the production of speech sounds," *J. Commun. Disord.* **39**, 350–365.
- Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). "Representation of sound categories in auditory cortical maps," *J. Speech Hear. Res.* **47**, 46–57.
- Hafke, H. Z. (2008). "Nonconscious control of fundamental voice frequency," *J. Acoust. Soc. Am.* **123**, 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.
- Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (2001). "Effects of delayed auditory feedback (DAF) on the pitch-shift reflex," *J. Acoust. Soc. Am.* **109**, 2146–2152.
- Hawco, C. S., and Jones, J. A. (2009). "Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals," *Brain Res.* **1276**, 131–139.
- Hirano, M., and Ohala, J. (1969). "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles," *J. Speech Hear. Res.* **12**, 362–373.
- Hirose, H., and Gay, T. (1972). "The activity of the intrinsic laryngeal muscles in voicing control," *Phonetica* **25**, 140–164.
- Horii, Y. (1984). "Phonatory initiation, termination, and vocal frequency change reaction times of stutterers," *J. Fluency Disord.* **9**, 115–124.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.
- Izdebski, K., and Shipp, T. (1978). "Minimal reaction times for phonatory initiation," *J. Speech Hear. Res.* **21**, 638–651.
- Jafari, M., Wong, K.-H., Behbehani, K., and Kondraske, G. V. (1989). "Performance characterization of human pitch control system: An acoustic approach," *J. Acoust. Soc. Am.* **85**, 1322–1328.
- Knornschild, M., Nagy, M., Metz, M., Mayer, F., and von Helversen, O. (2010). "Complex vocal imitation during ontogeny in a bat," *Biol. Lett.* **6**, 156–159.
- Koval, M. J., Lomber, S. G., and Everling, S. (2011). "Prefrontal cortex deactivation in macaques alters activity in the superior colliculus and impairs voluntary control of saccades," *J. Neurosci.* **31**, 8659–8668.
- Kuhl, P. K., and Meltzoff, A. N. (1996). "Infant vocalizations in response to speech: Vocal imitation and developmental change," *J. Acoust. Soc. Am.* **100**, 2425–2438.
- Larson, C. R. (1998). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control," *J. Commun. Disord.* **31**, 489–503.
- Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (2008). "Interactions between auditory and somatosensory feedback for voice F0 control," *Exp. Brain Res.* **187**, 613–621.
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (2001). "Comparisons of voice F0 responses to pitch-shift onset and offset conditions," *J. Acoust. Soc. Am.* **110**, 2845–2848.
- Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (2000). "Effects of pitch-shift onset velocity on voice F0 responses," *J. Acoust. Soc. Am.* **107**, 559–564.
- Larson, C. R., Sun, J., and Hain, T. C. (2007). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," *J. Acoust. Soc. Am.* **121**, 2862–2872.
- Leonard, R. J., Ringel, R., Horii, Y., and Daniloff, R. (1988). "Vocal shadowing in singers and nonsingers," *J. Speech Hear. Res.* **31**, 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.
- Liu, H., Behroozmand, R., Bove, M., and Larson, C. R. (2011). "Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback," *J. Acoust. Soc. Am.* **129**, 3946–3954.
- Liu, H., and Larson, C. R. (2007). "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," *J. Acoust. Soc. Am.* **122**, 3671–3677.
- Ludlow, C. L., and Lou, G. (1996). "Observations on human laryngeal muscle control," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular Publishing, San Diego), pp. 201–218.
- Musiek, F. E., Baran, J. A., and Pinheiro, M. L. (1990). "Duration pattern recognition in normal subjects and patients with cerebral and cochlear lesions," *Audiology* **29**, 304–313.
- Musiek, F. E., and Pinheiro, M. L. (1987). "Frequency patterns in cochlear, brain stem, and cerebral lesions," *Audiology* **26**, 79–88.
- 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.
- Nudelman, H. B., Herbrich, K. E., Hess, K. R., Hoyt, B. D., and Rosenfield, D. B. (1992). "A model of the phonatory response time of stutterers and fluent speakers to frequency-modulated tones," *J. Acoust. Soc. Am.* **92**, 1882–1888.
- 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**, 314–322.
- 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.
- Purcell, D. W., and Munhall, K. G. (2006). "Compensation following real-time manipulation of formants in isolated vowels," *J. Acoust. Soc. Am.* **119**, 2288–2297.
- Reiss, D., and McCowan, B. (1993). "Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning," *J. Comp. Psychol.* **107**, 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 truncatus*," *J. Comp. Psychol.* **98**, 10–28.
- Roubeau, B., Chevrie-Muller, C., and Guily, J. L. S. (1997). "Electromyographic activity of strap and cricothyroid muscles in pitch change," *Acta Otolaryngol.* **117**, 459–464.
- Schafer, R. J., and Moore, T. (2011). "Selective attention from voluntary control of neurons in prefrontal cortex," *Science* **332**, 1568–1571.
- Seidlits, S. K., Reza, T., Briand, K. A., and Sereno, A. B. (2003). "Voluntary spatial attention has different effects on voluntary and reflexive saccades," *Sci. World J.* **3**, 881–902.
- Shipp, T., Izdebski, K., and Morrissey, P. (1984). "Physiologic stages of vocal reaction time," *J. Speech Hear. Res.* **27**, 173–178.

- Shockley, K., Sabadini, L., and Fowler, C. A. (2004). "Imitation in shadowing words," *Percept. Psychophys.* **66**, 422–429.
- Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (2005). "Voice responses to changes in pitch of voice or tone auditory feedback," *J. Acoust. Soc. Am.* **117**, 850–857.
- Watson, B. C. (1994). "Foreperiod duration, range, and ordering effects on acoustic LRT in normal speakers," *J. Voice* **8**, 248–254.
- Wilbrecht, L., and Nottebohm, F. (2003). "Vocal learning in birds and humans," *Mental Retard. Dev. Disabil. Res. Rev.* **9**, 135–148.
- Zarate, J. M., and Zatorre, R. J. (2008). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," *Neuroimage* **40**, 1871–1887.
- Zollinger, S. A., and Suthers, R. A. (2004). "Motor mechanisms of a vocal mimic: Implications for birdsong production," *Proc. Biol. Sci.* **271**, 483–491.