Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control

Roosbeh Behroozmand, Oleg Korzyukov, Lindsey Sattler, and Charles R. Larson^{a)}

Speech Physiology Lab, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

(Received 7 November 2011; revised 23 July 2012; accepted 25 July 2012)

The present study describes a technique for analysis of vocal responses to auditory feedback pitch perturbations in which individual trials are first sorted according to response direction and then separately averaged in groups of upward or downward responses. In experiment 1, the stimulus direction was predictable (all upward) but magnitude was randomized between +100, +200, or +500 cents (unpredictable). Results showed that pitch-shift stimuli (PSS) of +100 and +200 cents elicited significantly larger opposing (compensatory) responses than +500 cent stimuli, but no such effect was observed for “following” responses. In experiment 2, subjects were tested in three blocks of trials where for the first two, PSS magnitude and direction were predictable (block 1+100 and block 2–100 cents), and in block 3, the magnitude was predictable (± 100 cents) but direction was randomized (upward or downward). Results showed there were slightly more opposing than following responses for predictable PSS direction, but randomized directions led to significantly more opposing than following responses. Results suggest that predictability of stimulus direction and magnitude can modulate vocal responses to feedback pitch perturbations. The function and causes of the opposing and following responses are unknown, but there may be two different neural mechanisms involved in their production. © 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4746984>]

PACS number(s): 43.64.Sj, 43.66.Hg, 43.70.Bk, 43.70.Jt [AL]

Pages: 2468–2477

I. INTRODUCTION

In recent years there has been a dramatic growth in the number of studies using the voice auditory feedback perturbation technique. While the earlier studies mainly dealt with voice pitch perturbations (Burnett *et al.*, 1998; Hain *et al.*, 2000; Jones and Munhall, 2000; Donath *et al.*, 2002; Natke *et al.*, 2002; Xu *et al.*, 2004; Bauer *et al.*, 2006; Chen *et al.*, 2007; Larson *et al.*, 2007), more recent studies have focused on shifting speech formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006; Munhall *et al.*, 2009). Both types of studies have yielded important information about the role of auditory feedback for the control of the voice during sustained vowel sounds, singing, and speech. During sustained vocalizations or speech, an upward shift in voice pitch feedback leads to a decrease in voice fundamental frequency (F0), and a decrease in pitch feedback leads to an increase in voice F0. The compensatory nature of these responses led to the suggestion that the audio-vocal system operates like a negative feedback control system that helps to stabilize voice F0 at an intended frequency. Shifting auditory feedback of voice formant frequencies also leads to the re-adjustment of speech articulators that compensates for formant perturbations, again suggesting a negative feedback control system that helps to control the acoustical properties of the voice during speech.

Another important property of the vocal responses to pitch-shifted voice feedback is that their onset latency is approximately 100 ms, suggesting that they are reflexive or automatic in nature. Such responses are most obvious for a short duration stimulus, e.g., 200 ms. However, for longer duration stimuli, >300 ms, two responses, or a single longer duration response may be observed (Burnett *et al.*, 1998). These responses were suggested to be due to voluntary mechanisms. In everyday vocal tasks, an early reflexive and later voluntary response would work together to stabilize the acoustical parameters of the voice. However, the reflexive component of a response rarely compensates completely for the stimulus magnitude. For example, a 100 cent stimulus (100 cents = 1 semitone) and 200 ms duration leads to a response of 20 to 50 cents, and considerable variability has been reported both within and between subjects (Burnett *et al.*, 1998). For longer stimulus durations, the magnitude is much larger and can equal the stimulus magnitude.

While most responses appear to oppose the stimulus direction, a few responses have been observed that change in the same direction as the stimulus, and these are termed following responses (Burnett *et al.*, 1998). Since these responses follow the stimulus, they would be disadvantageous for stabilizing voice F0, and in normal vocalizations could produce very abnormal voice frequencies. Early results indicated the following responses were observed most frequently with very large stimuli, e.g., 500 cents (Burnett *et al.*, 1998). It was suggested that these responses might be due to the subject treating the feedback signal itself as the referent, e.g., a piano note, and attempting to match it

^{a)}Author to whom correspondence should be addressed. Electronic mail: clarson@northwestern.edu

(Hain *et al.*, 2000). An alternative suggestion was that due to the nature of the stimulus, voice pitch, which perceptually consists both of a frequency and amplitude quality, the following responses may result from subjects mis-perceiving a change in voice pitch feedback and responding as though there was a change in voice loudness (Larson *et al.*, 2007). The result of this mis-perception could lead to the following responses.

The above studies used the signal “averaging” technique to measure vocal responses to perturbed auditory feedback because single-trial responses to a pitch-shift stimulus are often embedded in noise-like fluctuations of voice F0, and therefore, difficult to identify. The logic of this technique is that if a small response embedded in noise is time-locked to a stimulus, the noise contributions to the signal will average out to 0 across many stimuli, while the response linked to the stimulus will grow in comparison with the noise. It was reasoned that this same technique could be used for the study of vocal responses to brief acoustical signals (Sapir *et al.*, 1983). While there were no studies that quantified the number of feedback trials necessary to produce an acceptable averaged vocal response, Bauer and Larson (2003) determined that 15 to 25 trials either presented at a frequency of 1/vocalization or up to 5/vocalization, produced similar magnitude average responses. Several published studies since then have reported averages based on approximately 20 trials.

Despite the promising results from the study of the effects of perturbed auditory feedback on voice control described above, serious questions remain as to the correct methodology for conducting the studies. For example, there is high variability in response magnitudes, and averaged responses differ in their directions. Reducing response variability and understanding the cause of the differing response directions is critical if this technique is to be used to fully understand the role of auditory feedback on voice production. It is also important to understand that the signal averaging technique makes the assumption that most of the trials to be averaged consist of somewhat identical signals that reflect vocal reactions time-locked to the onset of the feedback perturbation stimuli. However, previous studies have shown that the individual trials are quite different; on some trials the signals oppose the stimulus direction and others follow it, and because of this, the traditional signal averaging technique may not be appropriate for the study of vocal responses to perturbations in the auditory feedback. Moreover, because the responses can differ in directions, it suggests there may be two different mechanisms underlying these responses.

An analogous sensory-motor system is that of muscle spindle proprioceptors that can be stretched or shortened, and they can respond with an increase or a decrease in activity (Matthews, 1972). In consideration of the fact that the pitch-shift reflex functionally behaves somewhat like the muscle spindle system, and must be able to respond in two directions (increase or decrease voice F0), the present study was undertaken to examine in detail the differential vocal responsiveness to pitch-shift stimuli under different conditions.

In this report, results of two experiments are presented in which a new method of signal averaging is used in which

individual trials were first sorted according to response direction (upwards or downwards) and then separately averaged according to their direction. The first experiment was done on vocal responses from a previously published study involving electro-encephalography (EEG) recordings in response to brief (200 ms) pitch-shift stimuli (PSS) in the middle of vocalization. The PSS had a predictable direction (all upward), but their magnitudes were randomized at +100, +200, and +500 cents (Behroozmand *et al.*, 2009). In that study, nearly 100 trials were recorded for each PSS magnitude (+100, +200, +500 cents), which allowed the examination of many more trials than what has been reported previously. A shortcoming of the study of Behroozmand *et al.* (2009) was that the anticipation of the predictable stimulus direction by the subjects may have influenced their vocal responses to the stimuli. Therefore, a second experiment was carried out that consisted of three separate blocks of trials in which the predictability of the stimulus direction was varied. In block 1, all stimuli were upwards (+100 cents), in block two they were downward (−100 cents), and in block 3, the PSS direction was randomized (+100 and −100 cents). This randomization procedure would remove any possible anticipation of stimulus direction on the part of the subjects. The subjects in experiment 2 were also tested in an ERP study, however, this report only deals with their vocal responses.

II. EXPERIMENT 1

A. Methods

1. Subjects

Fifteen right-handed native speakers of American English (10 females and 5 males, 18–27 years of age; mean age = 20.73 and std = 2.55) with no history of speech, hearing or neurological disorders, and no voice training participated in the study. All study procedures, including recruitment and data acquisition were approved by the Northwestern University institutional review board and subjects were monetarily compensated for their participation.

2. Procedures

The experiment consisted of three blocks of vocalization in which subjects maintained a steady vocalization of the vowel sound /a/ at their conversational pitch and loudness. During each vocalization trial, subjects were presented with eight pitch-shift stimuli (PSS) with 200 ms duration. The inter-stimulus interval (ISI) was randomized between 500–900 ms. Three different pitch-shift magnitudes (+100, +200, and +500 cents) were delivered randomly. During each vocalization block, subjects were asked to maintain the vowel sound for approximately 6–7 s and repeat this vocal task 25 times.

3. Instrumentation

Subjects were seated in a sound-treated room in which their voices were recorded with an AKG boomset microphone (model C420), amplified with a Mackie mixer (model 1202), and pitch-shifted through an Eventide Eclipse

Harmonizer. The pitch-shifted auditory feedback was heard through Etymotic earphones (model ER1-14A) inserted into the subject's ear canal. The gain between the subject's voice and feedback was further manipulated with a Crown amplifier (D75) to +10 dB sound pressure level, calibrated with a Zwislocki coupler and a Brüel & Kjær sound level meter (model 2250). The use of insert earphones along with the 10 dB gain between the voice and its feedback helped mask the airborne and bone conducted voice.

All parameters of the pitch-shifting auditory feedback stimulus, such as duration, magnitude, and ISI were controlled by MIDI software (MAX/MSP v.4.1. by Cycling 74). The MIDI software also generated a TTL pulse to mark the onset of each stimulus for synchronized averaging of vocal responses. Voice, feedback and TTL pulses were sampled at 10 kHz using PowerLab A/D Converter (Model ML880, AD Instruments) and recorded on a laboratory computer utilizing Chart software (AD Instruments).

4. Voice response analysis

The pitch frequency of the recorded voice signals was extracted in PRAAT using the autocorrelation method and then exported to MATLAB for further processing. The extracted pitch contours were segmented into single trial epochs ranging from −100 ms pre- and +500 ms post-stimulus and then converted from Hz to cents scale using the formula: $\text{cents} = 1200 \times \log_2(F_2/F_1)$, in which F_2 is a vector of pitch values for a single trial epoch and F_1 is the mean of pre-stimulus pitch from −100 to 0 ms (baseline).

5. Pre-sorting technique

The directional averaging technique described in this study entailed the sorting of individual trials into groups of upward and downward vocal responses to pitch-shifted voice feedback. The direction of responses in each individual trial was first determined by subtracting the mean amplitude of the voice F_0 contour in a 100 ms-long pre-stimulus window (−100 to 0 ms) from a post-stimulus window (50 to 250 ms). Because this technique requires a stable baseline upon which to perform the subtractive measures for determining response direction, an additional procedure was used to restrict data analysis to trials with low pre-stimulus variability. Specifically, the variability in the pre-stimulus window from −100 to 0 ms was measured, and if a component of the signal in that window exceeded a value of ± 15 cents, the trial was not included in the analysis. The value of 15 cents as a cut-off level was chosen based on the results of analysis using different values. Lower values than 15 cents led to rejection of too many trials, whereas values higher than 15 cents resulted in increased variability in the pre-stimulus waveform. Following the rejection of trials with excessive variability in the pre-stimulus window, upward and downward responses were defined by positive and negative values, respectively. Averaged responses were then calculated separately for upward and downward responses for each condition. The percentage of upward and downward trials for each condition was also calculated. Those responses that changed in

the downward direction were termed “opposing” responses because they opposed the direction of the pitch shift stimuli. Upward responses were termed “following” because they followed the stimulus direction. Response magnitudes and peak times were measured at the point of greatest deviation from the value at the onset of the response.

B. Results

The plots in Fig. 1 show the grand averaged vocal responses from 15 subjects to +100, +200, and +500 cents pitch shifts in voice feedback. The subplots in this figure show the time course of the following (upward), opposing (downward), and combined (opposing + following) vocal responses for each pitch-shift stimulus magnitude, separately. As can be seen both the upward and downward responses diverge from the combination of upward and downward averaged responses at latencies of about 30–40 ms after stimulus onset and reached a peak at about 200–250 ms. Figure 2 shows the same data plotted separately as a function of response direction, including the combined averages of upward and downward responses, for three different stimulus magnitudes. What is clear in this figure is that opposing responses to +100 and +200 cent stimuli are greater than those to +500 cent stimuli [Fig. 2(B)]. However, no such clear difference can be observed for the following responses to different pitch-shift stimulus magnitudes [Fig. 2(A)].

Repeated measures analysis of variances (Rm-ANOVAs) were used to test for significant differences between vocal response magnitudes (+100, +200, and +500 cents) in 10 ms increments starting after the stimulus onset. *Post hoc* tests were then performed using Bonferroni's correction method to test for the significance of pair-wise differences between vocal response magnitudes to 100 vs 200 cents, 100 vs 500 cents, and 200 vs 500 cents stimuli. For the opposing vocal responses, results of the analysis revealed a significant main effect of stimulus magnitude ($p < 0.05$) for time latencies starting from 290 ms [$F(2,28) = 3.91$, $p = 0.032$] to 390 ms [$F(2,28) = 3.51$, $p = 0.044$] post-stimulus. *Post hoc* tests revealed that the significant main effect within this time window was only due to significant ($p < 0.05$) differences between vocal responses to 100 vs 500 and 200 vs 500 cents stimuli, indicating that subjects produced significantly larger opposing responses (more negative) to 100 and 200 compared with 500 cents pitch shift stimuli in the 290 to 390 ms time window [marked with an asterisk above the opposing subplot in Fig. 2(B)]. Results showed no significant difference between opposing vocal responses to 100 vs 200 cents stimuli. Results of the analysis did not show any significant difference between following or combined (following + opposing) vocal responses for different stimulus magnitudes at any time.

The opposing responses reached a peak magnitude of −13.66 (std: 1.37), −13.64 (std: 1.12), and −10.36 (std: 0.49) cents at 258, 292, and 236 ms for +100, +200, and +500 stimuli, respectively. The following responses reached a peak magnitude of +14.35 (std: 1.44), +11.82 (std: 1.49), and +11.71 (std: 1.17) cents at 272, 249, and 288 ms for +100, +200, and +500 stimuli, respectively. A Rm-ANOVA test on the peak magnitudes of vocal responses only revealed a

FIG. 1. Graphs depicting voice F0 responses to pitch shift stimuli of +100 cents (A), +200 cents (B), and +500 cents (C). Solid black curves show responses changing in the opposite direction to the stimulus (opposing), dashed black lines show responses changing in the same direction as the stimulus (following), and grey lines show responses that are a combination of opposing and following responses. Vertical bars represent standard error of the mean of the averaged traces.

significant main effect of PSS magnitude [$F(2,28) = 3.42$, $p = 0.047$] for opposing responses. *Post hoc* tests using Bonferroni's correction showed that the significant PSS magnitude effect was due to a significantly smaller (less negative) opposing vocal responses to +500 cents pitch shifts compared with +100 cents ($p = 0.044$) and +200 cents ($p = 0.028$) stimuli [Fig. 2(B)].

The data in Table I show the mean of the total number of recorded vocal responses to +100, +200, and +500 cents stimuli for all 15 subjects. As can be seen in this table, subjects produced a large proportion (~50% or more on average) of following responses to feedback pitch perturbations. For +100 cents stimuli, out of 71 total trials, 48% were following response compared with 52% opposing responses. The proportion of opposing responses was slightly reduced

to approximately 48% and 43% for larger pitch shifts at +200 and +500 cents, compared with +100 cents stimulus, respectively. Results of the statistical analysis did not show any significant difference between the total numbers of opposing compared with following vocal responses across different PSS magnitudes.

III. EXPERIMENT 2

A. Subjects

Fifteen young adults (8 males and 7 females, mean age: 21) with normal hearing and no history of neurological or speech/language disorders signed informed consent (approved by the Northwestern University IRB) to participate in this study for monetary compensation.

FIG. 2. Graphs depicting voice F0 responses to pitch shift stimuli plotted individually for opposing (A), following (B), and combined (C) opposing plus following responses. Black dashed lines show responses to 100 cent stimuli, solid black lines show responses to 200 cent stimuli, and grey lines show responses to 500 cent stimuli. Horizontal array of asterisks in (B) show time period where responses to 100 and 200 cent stimuli were significantly larger than those to 500 cent stimuli.

TABLE I. Mean of the total number and percent of responses by direction for three different pitch-shift magnitudes across 15 subjects.

PSS magnitude	Total number of responses	Opposing responses	Following responses
+100 cents	71	37 (52.11%)	34 (47.89%)
+200 cents	70	34 (48.57%)	36 (51.43%)
+500 cents	80	35 (43.75%)	45 (56.25%)

B. Methods

1. Instrumentation and analysis

During the testing, the same equipment that was used in experiment 1 (headphones, microphone and other electronic instruments), data analysis and pre-sorting procedures were used in this experiment.

2. Procedures

There were three separate blocks of trials in this experiment. In each of the blocks, subjects sat in a sound treated booth and sustained a vowel (/a/) vocalization for approximately

4–5 s, before pausing for a repetition. In the UP block of trials, all pitch-shift stimuli (200 ms duration) were in the upward direction, in the DOWN block all stimuli were in the downward direction and in the RANDOM block, stimulus directions were randomly varied in either direction. Thus in the first two conditions, subjects would realize that all stimuli were in the same direction, which might have caused them to form a prediction about the upcoming stimuli and influenced their response. However, in the third condition with the randomized stimulus directions, it would not be possible to predict the stimulus direction. For the UP and DOWN blocks, subjects vocalized 50 times with three pitch-shift stimuli per trial, and therefore received 150 pitch-shift stimuli in each block. In the RANDOM block, subjects vocalized 100 times, receiving three pitch-shifts per trial, resulting in 150 pitch-shift stimuli in each direction. The stimulus onset time varied from 500 to 1000 ms following vocal onset, and the inter-stimulus interval was 600–800 ms.

C. Results

As with experiment 1, all subjects produced both opposing and following responses. Figure 3 presents grand averaged

FIG. 3. Graphs depicting voice F0 responses to pitch shift stimuli in the upward direction (left) and downward direction (right). Top row of plots show responses to stimuli that were randomized in their direction (unpredictable direction). Bottom row of plots show responses to stimuli that were not randomized in their direction (predictable direction). Solid black traces show opposing responses, black dashed traces show following responses, and grey traces show the combination of opposing and following responses. Arrows indicate time of marked change in rate of return of following responses towards baseline discussed in text. Square brackets at bottom show time and direction of stimuli.

responses across all subjects for the three conditions, UP not random (predictable), DOWN not random (predictable), and RANDOM unpredictable stimulus directions. The upper two plots show responses for the randomized (unpredictable) condition, and the bottom two show responses for non-randomized (predictable) stimuli. The left column shows responses to upward stimuli, and the right column to downward stimuli. The grey traces in the middle of each plot show grand averages of combined upward and downward responses, i.e., using the conventional method of averaging. Above and below these traces are averages of responses in the opposing direction (black traces) and following direction (dashed traces). It is apparent from these graphs that the magnitude of the combined grand averages is less than the opposing or following averages, which is due to the fact that in the process of averaging, upward and downward responses tended to cancel each other when they are combined.

As has been shown in previous studies using this pitch perturbation technique, there was considerable variability from subject to subject. Figure 4 shows bar plots of the response magnitudes for upward and downward responses across subjects using the pre-sorting procedure. With the pre-sorting technique, there is still a fairly wide distribution of response magnitudes between subjects as has been reported previously (Burnett *et al.*, 1998; Munhall *et al.*, 2009).

Separate three-way repeated measures ANOVAs were done to test the significance of the timing and magnitude of the response peaks. Results of the Rm-ANOVA for response magnitude by response type (opposing vs following), direction (up or down) and predictability (randomized vs non-randomized) failed to show a significant effect for predictability [$F(1,14)=0.281$, $p>0.05$] or direction [$F(1,14)=0.355$, $p>0.05$], however, there was a significant effect for response type [$F(1,14)=4.78$, $p<0.05$]. Opposing response type magnitudes (mean: 24.03) were significantly larger than following responses (mean: 20.55 cents). There was also a significant interaction between response type and predictability [$F(1,14)=9.12$, $p<0.009$]. *Post hoc* t-tests showed that the opposing responses to unpredictable downward stimuli had a significantly ($p<0.01$) larger magnitude

FIG. 4. Bar plots of mean response magnitude for upward and downward responses for the pre-sorted data. Plots illustrate gradation of response magnitudes by subject and arranged from smallest to largest magnitudes on the x-axis.

TABLE II. Mean peak times (SD) of opposing and following responses for the randomized vs non-randomized stimulus directions.

	Random	Non-random
Opposing	318.2 ms (74)	308.2 ms (95)
Following	203.5.9 ms (95)	207.4 ms (84)

(mean: 25.06 cents) than the following responses (mean: 19.57 cents) to unpredictable downward stimuli. Furthermore, the opposing responses to unpredictable upward stimuli were significantly ($p<0.0001$) larger in magnitude (mean: 24.39 cents) compared with the following responses to upward stimuli in the unpredictable condition (mean: 17.38 cents).

The three-way Rm-ANOVA testing of response peak time by response type, direction and predictability revealed a significant difference for response type [$F(1,14)=28.19$, $p<0.0001$]. Following responses had a significantly earlier peak time (205, SD 88 ms) than opposing responses (314, SD 84 ms). There was no significant difference for direction [$F(1,14)=0.985$, $p>0.05$] or predictability [$F(1,14)=0.044$, $p>0.05$], and there were no significant interactions. Table II presents the peak (maximum or minimum) time of the response for both opposing and following responses for the random and non-random conditions.

The differences in the peak times of the responses are clearly seen in Fig. 3. Also, an important feature to note in these graphs is that the F0 contour of the following responses reached a peak (or trough) at about 166 ms and then reversed direction to head back towards the baseline running parallel to the slope of the opposing response average. At the inflection point indicated by the arrows (approximately 300 ms), the rate of change was reduced but still moving towards the baseline. Thus, the initial part of the following response was in the same direction as the stimulus, and the later part was in the opposite direction to the stimulus. The F0 contours of the opposing responses had a somewhat slower onset rate of change, reached a peak at about 300 ms, and then reversed direction to head back towards the baseline.

Statistical testing for percentages of response type was only done for the opposing responses, since the following response percentages are simply the inverse of opposing response percentages. A two-way Rm-ANOVA test of the percentage of opposing responses by response direction and predictability failed to show significant differences for direction [$F(1,29)=0.017$, $p>0.05$] or for predictability [$F(1,28)=1.21$, $p>0.05$].

To test the effect of the sorting procedure against the non-sorting (traditional analysis technique), a four-way ANOVA was run for response magnitude by the sorting condition, response type (opposing vs following), response direction (up vs down) and randomization. Results showed that sorting responses led to larger response magnitudes (22 cents) than the non-sorting method (13 cents) [$F(1,175)=27.9$, $p<0.0001$]. None of the other tested effects, including stimulus direction, differed significantly. Repeated measures ANOVA could not be run on these measures because of unequal numbers of measures in each condition.

IV. DISCUSSION

The present study proposed the directional averaging technique as a new method for improving the analysis of vocal responses to pitch perturbations in voice auditory feedback. Previous studies have shown high levels of variability in vocal responses to pitch-shift stimuli both within and between subjects (Burnett *et al.*, 1998), which detracts from the potential usefulness of the technique for studying the underlying mechanisms of feedback-based control of vocal output. While there is still large variability in response magnitudes between subjects, the responses are uniformly larger with the pre-sorting technique compared to the traditional non-sorting technique. The present technique was developed to help reduce response variability and to improve our understanding of the differences between the opposing and following responses. Pre-sorting responses by their direction and then averaging responses of like direction, allowed us to show that the traditional averaging method is flawed because it mixes responses of different directions and therefore does not accurately portray the true nature of the opposing and following responses. Moreover, it appears that the prevalence of opposing and following responses is nearly equal in some cases.

Previous studies have demonstrated that there are two different types of responses (opposing and following) to pitch perturbation in voice feedback, suggesting that the reflexive vocal reactions to changes in voice pitch auditory feedback may consist of two distinctly different neural circuitries that serve different functions for online monitoring and control of voice F0 (Hain *et al.*, 2000). The opposing responses were suggested to be corrective in nature and stabilize vocal pitch output near an intended frequency by changing voice F0 in the opposite direction to a perturbation in voice pitch feedback. In contrast, the following responses tend to de-stabilize vocal output by changing voice F0 in the same direction as the perturbation in voice pitch feedback. Although a satisfactory explanation of the following responses has never been determined, it has been suggested that they occur in conditions when the audio-vocal system treats the altered feedback as an external referent (e.g., a piano note) and attempts to match it by changing vocal output to follow the feedback (Hain *et al.*, 2000). Results of the present study raise questions about this interpretation of the responses and lead to a re-evaluation of the underlying mechanisms and suggestions for improved methodology in future studies.

The method of sorting responses by direction prior to averaging used in this study was developed after consideration of alternative techniques, such as a post-stimulus threshold technique. It was determined that of all techniques tried, some have their advantages and disadvantages. Moreover, the windowing dimensions used in the present study, 100 ms before the stimulus and 50 to 250 ms following the stimulus had its flaws also. These window dimensions captured the predominant direction of most responses, however, it was noted that no matter which window dimensions were used, there were always some responses that seemed questionable as to whether they should be following or opposing. These

questionable cases were a result of pre- and post-stimulus variability. In general, out of all the windowing dimensions tested, within a given subject, the relative numbers of opposing and following responses were fairly consistent. Thus, it seems that the crucial factor in establishing windows for comparing baseline with the post-stimulus response is that the post-stimulus window should be centered over the time when most response peaks occur, 200–300 ms, and a pre-stimulus window that reflects the mean of the pre-stimulus baseline.

A. Response magnitudes

An important observation of experiment 1 was the difference in response magnitudes between opposing and following responses for stimuli of 100 and 200 cent magnitudes and those observed for 500 cent stimuli. The magnitude of compensatory (opposing) vocal responses were significantly larger (more negative) for smaller (i.e., +100 and +200 cents) compared with larger (i.e., +500 cents) stimuli in a time window spanning 290–390 ms after stimulus onset [see Fig. 2(B)]. However, no significant effect of stimulus magnitude was found for the magnitude of following or combined (following + opposing) vocal responses. These findings suggest that smaller pitch-shift magnitudes can elicit larger compensatory vocal responses than do larger stimuli. This finding along with data from a previous study (Liu and Larson, 2007) provide converging evidence that the feedback-based control of vocal parameters is more effective if there is a small discrepancy between the intended vocal output and its feedback. One explanation for this effect is that the audio-vocal system may utilize the efference copies of motor commands (Wolpert *et al.*, 1995; Wolpert, 1997) to estimate the degree of discrepancy between intended (predicted) and actual (sensory) feedback in order to identify the source of sensory stimulation. Those inputs that closely match the intended output are interpreted as being self-produced whereas those that have large discrepancies are interpreted to be generated by an external source. Therefore, it is possible that smaller perturbations (+100 and +200 cents) in the current study elicit larger vocal responses compared to +500 cents stimuli because the system recognizes them as being self-generated and attempts to counteract them by generating larger compensatory (opposing) responses than for larger magnitude stimuli. However, it should be noted that previous studies did not show significantly smaller responses for 500 ms stimuli compared with 100 or 200 cent stimuli, possibly because previous studies mixed both types of responses in the averaging process (Burnett *et al.*, 1998). The results here also show it is desirable to average more responses than were typically done previously so as to be able to get an adequate number of both types of responses in the same condition.

In contrast with opposing responses, the following responses were not affected by the stimulus magnitude; following response magnitudes were nearly identical for +100, +200, and +500 stimuli [Fig. 2(A)]. Since the following responses are de-stabilizing in nature (positive feedback mode), they tend to increase the vocal pitch error in the

same direction as changes in the auditory feedback. This finding suggests the neural mechanisms controlling the following responses are different than those controlling the opposing responses. Moreover, these differences are not just a difference in direction; the following responses do not appear to be modulated by efference mechanisms as has been suggested for opposing responses (Hain *et al.*, 2000), possibly because the former do not function as part of a negative feedback system that helps to stabilize voice F0. If the following responses do not serve this function, then they may not be controlled by the same circuitry that is involved in discriminating between self and non-self-vocalizations.

In experiment 2, there were no significant differences in response magnitudes as a function of predictability, stimulus direction or type of response. In a previous study (Liu *et al.*, 2011a) it was observed that response magnitudes to downward stimuli were larger than those to upward stimuli. Results from the present study, experiment 2, RANDOM condition, failed to support this finding. One possible explanation for this difference is that in the Liu *et al.* study, during each vocalization, all stimuli were shifted six times in the same direction. In the present study, RANDOM condition, the three stimulus directions were randomized even within a single vocalization. Thus, during a single vocalization if all stimuli were in the same direction, subjects may have formed a prediction about the stimulus, which led to smaller response magnitudes. However, this explanation does not explain why the responses to downward responses were larger than those to upward stimuli in the Liu *et al.* study. Clearly, this issue needs further investigation.

B. Percent of response types

Another important finding of the present study was that in conditions where the stimulus was always in the same direction (predictable), the number of opposing and following responses was nearly equal (see Table I, experiment 1 and Table II, experiment 2). For smaller stimulus magnitudes (i.e., 100 cents), subjects produced a slightly higher percent of opposing (~52%) compared with following (~48%) responses. However, for larger stimulus magnitudes (i.e., +200 and 500 cents) in experiment 1, there was a slight increase in the relative percentage of following responses compared with the opposing ones (~51% following vs ~49% opposing for +200 cents and ~56% following vs ~44% opposing for +500 cents stimuli), supporting a previous observation by Burnett *et al.* (1998) in which a larger proportion of following responses were observed for larger magnitude stimuli.

The near equal numbers of opposing and following responses in Exp.1 and the non-random (predictable) condition of experiment 2 suggests that the neural mechanisms involved in detecting the stimulus direction and producing the appropriate responses are sensitive either to the predictability of the stimulus direction or that the stimuli are all in the same direction. Thus, if all stimuli are in the same direction and therefore predictable, subjects may develop a propensity to respond in the same way whenever there is a change in feedback. One possible explanation for this behavior

may relate to vocal imitation. There is solid evidence for the fact that other animals and human infants learn vocalizations and speech by imitation of what they hear (Davis *et al.*, 1993; Greenlee *et al.*, 2011; Liu *et al.*, 2011b). Since learning by imitation is common among different animal species, this ability may be conserved through evolution and naturally lead to the tendency to imitate what is heard. In the context of the experiments described here, as a subject hears his/her voice repetitively change in the same manner across many trials, they may start imitating the sound of their voice changing in the same manner to each stimulus.

C. Response timing

An important difference between the results of experiments 1 and 2 is in the peak time of the following responses relative to the opposing response. The earlier peak time seen in experiment 2 (Fig. 3) is also coupled with a reversal in direction, with the latter part of the response moving in the opposing direction to the first part. From Fig. 1, it may be seen that the averages of some of the following responses resemble those in Fig. 3, although the shape of the differences with the opposing responses are not as distinct. The only methodological differences in the two experiments were the randomization of the stimulus magnitude and direction, and it is difficult to understand why these variables would affect the shape of the responses. A possible explanation for these differences in shape is that both types of response begin as a continuation of voice F0 tremor that exists prior to the stimulus, and at the onset of the stimulus, the F0 tremor continues as a rise or fall in F0 at stimulus onset. Then after about 40 to 50 ms the neural triggered response begins (Liu *et al.*, 2011a). Thus, the following responses begin at stimulus onset as a fluctuation in F0 that continues in the same direction as the stimulus, but then the neuromuscular reflexive mechanisms slow the F0 trajectory, cause it to reach a peak at about 166 ms, and reverse it to the opposing direction. These responses appear to end at about 300 ms (arrows in Fig. 3), which is about the same time that the opposing responses reach their peak and begin to return to baseline. When the opposing responses are triggered, they continue in their direction away from the baseline and opposite to the stimulus. Thus, it is suggested that many following responses start as a random fluctuation in F0 at stimulus onset and then are reversed in direction by the neuromuscular mechanisms that terminate them and cause a return to baseline. However, it is still possible that there are other following responses that might be “true following” responses, i.e., those that are driven by neural mechanisms that attempt to change voice F0 to the same frequency as the pitch-shift stimulus, but due to the nature of the variability in individual responses and the resulting ambiguity in the averaging process, it is difficult at this time to say more about them.

The above explanation for the following responses may also explain the near equal numbers of opposing and following responses when all the stimuli are in the same direction (experiment 1), and in the UP or DOWN experiments of experiment 2. If indeed many responses begin as a fluctuation in vocal tremor, the phase of the tremor should be

randomized with respect to the stimulus; therefore, it may be assumed that on 50% of the trials, the tremor will be in the opposing direction and on 50% to the following direction of the stimulus. Indeed, the near 50% distribution of both response types described in experiment 1 could be explained by the randomized phase relations of the two types of responses.

The larger question arising from the experiments when *all* stimuli are in the *same direction* is why there were not more opposing responses, since it might be predicted that the neural mechanisms would favor a corrective response to a perturbing stimulus. One possible explanation is that the neural mechanisms governing the responses adapt to the same stimulus direction occurring over and over again, and become less sensitive to the stimuli. Again, if subjects have a tendency to imitate vocal sounds or a pattern of sounds, the repetition of the same sound over and over again, may be a factor leading to the high percentage of following responses in the non-random condition.

If subjects habituate to repetitive stimuli of the same direction, it suggests that EEG or fMRI experiments conducted using the pitch-shift paradigm where only a single response direction is used, may not accurately reflect the neural mechanisms involved in the detection of “unexpected” perturbations. This possibility also leads to the suggestion that it is important to randomly vary the stimulus directions in order to more accurately assess the underlying neural mechanisms.

Along with this above suggestion, as shown in Figs. 1 and 3, when analysis of the vocal responses does not take into consideration the direction of the stimuli, the addition of upward and downward responses leads to very small averaged response magnitudes, which no doubt is caused by cancellation due to the addition of positive and negative values. This cancellation of different direction responses may account for much of the variability in response magnitudes in previous studies. In order to circumvent this possibility, it would be best to pre-sort the responses prior to signal averaging. Such pre-sorting should be done not only on the vocal response but also on EEG or fMRI signals in studies of the neural mechanisms underlying the responses.

The cancellation may also affect the response latencies. As seen in Fig. 1 where the separation of the opposing and following traces begin shortly after the stimulus, the trace representing the addition of the responses (middle trace) does not show a response onset until slightly later. Shortly after the stimulus, the beginning of the upwards and downwards responses, which are no doubt quite small, easily cancel each other, and it is not until slightly later in time that larger or more numerous response types predominate and give rise to one or the other types of responses. Moreover, as discussed above, if the responses begin as a variation in voice tremor, any measure of the exact time when the response begins is likely to be in error. It is for this reason that response latencies have not been presented in this paper. It is suggested that a much more easily defined measure of response timing be made at the time of the greatest magnitude of the response (peak time). In the case of Fig. 1, these times are between 200 and 300 ms. Describing response timing according to this metric is similar to the way ERP (EEG) responses are defined.

In experiment 2, stimulus magnitudes were kept constant, but the predictability of the stimulus direction affected the magnitudes. Opposing responses had larger magnitudes when the stimulus direction was unpredictable, but following responses were not affected by the predictability. These differences when combined with the results of experiment 1 tell us that unexpected pitch-shift stimuli trigger responses that are a reaction to the stimulus itself, i.e., its magnitude and direction. In contrast, the following responses do not appear to be closely linked to the stimuli. The question then becomes what do the following responses represent and how are they modulated?

Finally, in previous studies, one of the primary dependent variables used to quantify the relation between pitch-shift stimuli, and the vocal response, was the response magnitude. Although, this measure is still a justified measure, the present study suggests that the measure of “percentage of response types,” i.e., following or opposing may be another useful measure to characterize neural mechanisms of vocal control based on feedback perturbations. There was wide variability in this measure in the subjects tested in the two experiments described here. What is not known is how the percentage of response type within a group of subjects relates to neural mechanisms of vocal motor control. That is, a high percentage of opposing responses may indicate that subjects rely heavily on auditory feedback for vocal control, whereas a high percentage of following responses may indicate a lack of attention to auditory feedback in vocal control. Such lack of attention may be related to poor vocal training or poor voice control. These speculations should be tested in groups of subjects with known vocal control characteristics in order to verify whether or not this metric might be useful for improving our understanding of neural mechanisms of voice control.

V. CONCLUSION

In summary, the present study describes a new technique for averaging voice F0 responses to pitch-shifted voice feedback. This technique pre-sorts individual trials on the basis of their direction at the onset of the stimulus and then separately averages upward and downward responses. This technique reveals that there are many more following responses than previously reported and the magnitudes of the opposing and following responses are about the same. It was also shown (experiment 1) that opposing response magnitudes are larger for smaller stimuli (+100 and +200 cents) than larger stimuli (+500 cents), whereas magnitudes of following responses did not differ with stimulus magnitudes. The latter findings support the idea that smaller pitch-shift stimuli are more likely to be recognized as self-produced and therefore they represent an error in production requiring a compensatory, correcting response. In contrast, the insensitivity of following responses to stimulus magnitude suggests they may not be involved in stabilizing voice F0 as was suggested for the opposing responses (Hain *et al.*, 2000). All together, the present study has demonstrated that the traditional method of averaging vocal responses to pitch-shifted feedback has an inherent flaw that distorts the numbers of

opposing and following responses that have been reported in previous studies.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. 1R01 DC 006243. We wish to thank Mr. Chun Liang Chan for his assistance in programming.

- Bauer, J. J., and Larson, C. R. (2003). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," *J. Acoust. Soc. Am.* **114**, 1048–1054.
- Bauer, J. J., 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**, 2363–2371.
- 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.
- 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.
- 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.
- Davis, P. J., Zhang, S. P., and Bandler, R. (1993). "Pulmonary and upper airway afferent influences on the motor pattern of vocalization evoked by excitation of the midbrain periaqueductal gray of the cat," *Brain Res.* **607**, 61–80.
- 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., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Hiroto, K., Chen, H., and Howard, M. A., 3rd (2011). "Human auditory cortical activation during self-vocalization," *PLOS One* **6**, 1–15.
- 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.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F0 production: Evidence from feedback perturbation," *J. Acoust. Soc. Am.* **108**, 1246–1251.
- 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.
- Liu, H., Behroozmand, R., Bove, M., and Larson, C. R. (2011a). "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.
- Liu, H., Meshman, M., Behroozmand, R., and Larson, C. R. (2011b). "Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback," *Clin. Neurophysiol.* **122**, 951–957.
- Matthews, P. B. C. (1972). *Mammalian Muscle Receptors and Their Central Actions* (Williams & Wilkins, Baltimore), pp. 1–630.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. S. (2009). "Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate," *J. Acoust. Soc. Am.* **125**, 384–390.
- Natke, U., Grosser, J., Sandrieser, P., and Kalveram, K. T. (2002). "The duration component of the stress effect in stuttering," *J. Fluency Disord.* **27**, 305–317.
- 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.
- Sapir, S., McClean, M. D., and Larson, C. R. (1983). "Human laryngeal responses to auditory stimulation," *J. Acoust. Soc. Am.* **73**, 315–321.
- Wolpert, D. M. (1997). "Computational approaches to motor control," *Trends Cogn. Sci.* **1**, 209–216.
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). "An internal model for sensorimotor integration," *Science* **269**, 1880–1882.
- 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**, 1168–1178.