Published in final edited form as:

J Acoust Soc Am. 2003 August ; 114(2): 1048–1054.

Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique^{a)}

Jay J. Bauer^{b)} and Charles R. Larson

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

Abstract

The pitch-shift reflex is a sophisticated system that produces a “compensatory” response in voice F_0 that is opposite in direction to a change in voice pitch feedback (pitch-shift stimulus), thus correcting for the discrepancy between the intended voice F_0 and the feedback pitch. In order to more fully exploit the pitch-shift reflex as a tool for studying the influence of sensory feedback mechanisms underlying voice control, the optimal characteristics of the pitch-shift stimulus must be understood. The present study was undertaken to assess the effects of altering the duration of the interstimulus interval (ISI) and the number of trials comprising an average on measures of the pitch-shift reflex. Pitch-shift stimuli were presented to vocalizing subjects with ISI of 5.0, 2.5, 1.0, and 0.5 s to determine if an increase in ISI altered response properties. With each ISI, measures of event-related averages of the first 10, 15, 20, or 30 pitch-shift reflex responses were compared to see if increases in the number of responses comprising an event-related average altered response properties. Measures of response latency, peak time, magnitude, and prevalence were obtained for all ISI and average conditions. While quantitative measures were similar across ISI and averaging conditions, we observed more instances of “non-responses” with averages of ten trials as well as at an ISI of 0.5 s. These findings suggest an ISI of 1.0 s and an average consisting of at least 15 trials produce optimal results. Future studies using these stimulus parameters may produce more reliable data due to the fivefold decrease in subject participation time and a concomitant decrease in fatigue, boredom, and inattention.

I. INTRODUCTION

Neural mechanisms controlling vocalization are poorly understood. Recent years have seen increased study of mechanisms of vocal control in primates and birds (Jürgens, 2002; Luthe *et al.*, 2000; Solls *et al.*, 2000; Suthers *et al.*, 2002). While these studies have increased our understanding of voice control, their invasive nature has precluded parallel studies in humans. An alternative approach for studying vocal control mechanisms in human subjects is to analyze real-time physiologic responses to unanticipated perturbations in sensory feedback during ongoing motor tasks. Such perturbations can mimic naturally occurring sensory events that arise from execution of controlled motor tasks, and can reveal important properties of the underlying neural control mechanisms. For example, in the study of mechanisms controlling orofacial musculature during speech, systematic application of mechanical loads to lips during speech demonstrates that the nervous system uses sensory information from the lips for predictive,

^{a)}Material originally presented in “Improvements in methodology for the pitch-shifting technique,” proceedings of The Acoustical Society of America, Chicago, IL, June 2001.

^{b)}Electronic mail: j-bauer2@northwestern.edu.

feed-forward purposes (Abbs and Gracco, 1984; Gracco, 1995; Saltzman *et al.*, 1998; Shaiman and Gracco, 2002).

The perturbation approach is also used to study the properties of the audio-vocal system. These investigations typically involve presenting pitch modulated auditory feedback to a vocalizing subject (Burnett *et al.*, 1998). During sustained vowel phonations and glissandos, the audio-vocal system operates in a negative feedback mode that serves to stabilize voice fundamental frequency (F_0) around the intended pitch (Burnett *et al.*, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000; Larson *et al.*, 2000). Thus, an automatic “compensatory” change in voice F_0 corrects for a discrepancy between the intended voice F_0 and the feedback pitch. Because of the automatic nature of the response and lack of habituation, the audio-vocal response to the unexpected change in pitch is termed the pitch-shift reflex (PSR) (Burnett *et al.*, 1998). The latency of the response to altered pitch feedback is ~100 to 150 ms (Hain *et al.*, 2001; Larson *et al.*, 2000), which suggests that underlying neural mechanisms are more complex than a simple multisynaptic reflex loop contained in the medulla. Rather, the pathways may involve higher levels of the brainstem, such as the midbrain periaqueductal gray (PAG) and inferior colliculus (IC), or even cortical and cerebellar pathways. Recent research indicates that the pitch-shift reflex may also be used in control of F_0 related to suprasegmental aspects of speech (Donath *et al.*, 2002; Natke and Kalveram, 2001). Exploring the characteristics of the pitch-shift reflex may provide important information about how the nervous system uses auditory feedback in the regulation of voice F_0 during speech and singing.

If the pitch-shift reflex is to be useful for the study of the influence of auditory feedback on voice F_0 control, many parameters of the reflex must be understood. One parameter that has not been addressed in previous studies is interstimulus interval (ISI). Previous investigations of the pitch-shift reflex randomly presented one pitch-shift stimulus during the first few seconds of a sustained vowel repeated for 15–20 consecutive vocalizations. This paradigm corresponded to an interstimulus interval (ISI) greater than 5 s (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2000), which likely provided more than adequate recovery time between successive pitch perturbations. However, it remains unclear whether the audio-vocal system can respond to multiple pitch perturbations presented across the duration of a single vocalization. Stimulation at high rates may result in an overlap of successive responses if response latency and duration exceed the interval between consecutive stimuli. Such an overlap can introduce error, and alter the magnitude (Nelson and Lassman, 1968; Wikström *et al.*, 1996) of the observed response. Complex systems such as the pitch-shift reflex require time to process and transmit neural potentials from one location to another. Thus, temporal processing constraints limit the frequency at which the system can respond to repetitive stimulation without a reduction in response magnitude. We hypothesize that the audio-vocal system continuously monitors auditory feedback pitch throughout the duration of an utterance, and responds to repetitive pitch-shift stimulation with successive compensatory pitch-shift reflexes. We tested this hypothesis by comparing PSR responses elicited under several interstimulus intervals (ISI).

Event-related averaging techniques have been widely used to investigate small biological signals in a noisy environment. In studies such as auditory evoked potentials, thousands of averages are obtained with relative ease because subjects do not need to attend to the stimuli or actively control muscular movements (Abbas and Brown, 1991; Dawson, 1954; Nakamura *et al.*, 1989; Nelson and Lassman, 1968). However, in studies of motor systems, subjects must maintain attention, and control muscles within rather strict limits (Barlow and Bradford, 1996; Burnett *et al.*, 1998; Sapir and McClean, 1981). These demands can be taxing on subjects, but such limitations are partially offset by the fact that motor responses are typically large and reliable. Thus, averaged motor responses can be obtained from a noisy background with far fewer stimulus presentations than those needed to discern evoked brain responses. Studies of

speech motor systems have utilized event-related averaging techniques to study the role of sensory feedback on perioral, jaw, and laryngeal muscles (Barlow and Bradford, 1996; Burnett *et al.*, 1998; Sapir and McClean, 1981; Smith *et al.*, 1987). However, in these studies, the number of trials comprising each average varied from 15 to 80 trials. If an objective assessment of the number of averages is not known, experimenters may average insufficient trials to obtain an event-related average, or they may have used too many trials, which confers no greater response reliability and yet may excessively tire the subject. Nevertheless, it is important to understand interactions between stimulus and subject-dependent variables in the study of motor responses to sensory stimulation. Therefore, a secondary objective of the present study was to determine the optimal number of trials required for reliable averaged measurements of the pitch-shift reflex.

Determining the optimal stimulus parameters needed to elicit a pitch-shift reflex has the potential to produce more reliable data. A concern with our previous technique of presenting a single stimulus for each 5-s vocalization is that a lengthy period of vocalization was required in order to study several manipulations of an independent variable. Such extended vocalization time could lead to fatigue, boredom, and more variable data. In this study, we presented pitch-shift stimuli to vocalizing subjects with interstimulus intervals of 5.0, 2.5, 1.0, and 0.5 s. These intervals correspond to stimulation rates of 0.2, 0.5, 1.0, and 2.0 pitch-shifts per second. We then compared dependent measures across event-related averages consisting of the first 10, 15, 20, and 30 trials obtained at each ISI.

II. METHODS

Thirty-one normal-hearing young adult subjects (8 males and 23 females) with no history of speech-language disorders or neurological deficits were tested under the experimental conditions described below. In each condition, subjects repeatedly produced /u/ vowel sounds for ~5 s into an AKG boom-set microphone (model HSC 200) at a conversational voice F_0 at 70 dB SPL at a microphone-to-mouth distance of 5 cm, while seated comfortably in a sound attenuated booth. The microphone (voice) signal was amplified with a Mackie Mixer (model 1202), then processed for pitch-shifting with an Eventide Ultraharmonizer (SE 3000), mixed (Mackie Mixer model 1202-VLZ) with pink masking noise (Goldline Audio Noise Source, model PN2, spectral frequencies 1 to 5000 Hz), and presented to the subject over AKG headphones (model HSC 200) after amplification by a Crown amplifier (D75-A). Voice signals were amplified to 80 dB SPL and pink noise was amplified to 60 dB SPL to partially mask the detection of bone-conducted signals. The Ultraharmonizer was controlled by MIDI software to repeatedly insert pitch-shift stimuli of 0.1-s duration and 100 cents (100 cents=1 semitone) greater than speaking F_0 into the subject's vocal auditory feedback during each vocalization. These stimulus parameters were chosen because they have been shown to elicit reliable reflexes (Burnett *et al.*, 1998). Moreover, stimuli of 0.1-s duration do not elicit secondary, voluntary responses (Burnett *et al.*, 1998). Stimuli were presented at an average ISI of approximately 2.5, 1.0, and 0.5 s. Each ISI varied somewhat (~100 ms) to reduce the ability of subjects to predict the exact time of stimulus onset. Thirty stimuli were presented under each ISI condition. Thus, the number of vocalizations needed under each ISI condition differed. We collected 15 vocalizations for the 2.5-s ISI condition, 6 vocalizations for the 1.0-s ISI condition, and 3 vocalizations for the 0.5-s ISI condition. The order of presentation of the experimental conditions was randomized across subjects.

We tested ten of the original subjects (three males and seven females) under one additional ISI condition. Under this additional condition, pitch-shift stimuli were presented approximately every 5.0 s resulting in only one stimulus per vocalization across a total of 30 vocalizations. This ISI condition was included *posthoc* so that we could directly compare PSR responses elicited using repetitive pitch-shifts per vocalization with those elicited by a single pitch-shift

stimulus per vocalization, as in previous studies (Burnett *et al.*, 1998; Hain *et al.*, 2001, 2000; Larson *et al.*, 2000).

Each voice signal, feedback signal, and TTL pulse indicating the timing of the MIDI signal to the Ultraharmonizer and resultant pitch-shift was digitized on-line at 10 kHz (5 kHz anti-aliasing filter) onto a laboratory computer. In offline analysis in preparation for frequency extraction, the voice signal was low-pass filtered at 200 Hz for females and 100 Hz for males, differentiated, and then smoothed with a five-point binomial, sliding window (Larson *et al.*, 2000). Low-pass filtering of the voice signal removed most of the energy present at harmonic frequencies, allowing error-free triggering of the F_0 in subsequent stages of signal processing. A software algorithm then detected positive-going threshold-voltage crossings, interpolated the time fraction between each pair of sample points that constituted a crossing, and calculated the reciprocal of the period defined by the center points to signify the voice F_0 . The resulting F_0 signal was then low-pass filtered at 10 Hz to remove sharp discontinuities associated with each glottal cycle. For each rate condition, the F_0 signal was time aligned to stimulus onset (TTL pulse) and averaged. Event-related averages were calculated using a 0.2-s pretrigger baseline and 0.5-s posttrigger window. For each subject, separate averages were made for voice F_0 responses for the first 10, 15, 20, and 30 stimuli presented under each ISI condition.

A software algorithm was used to extract poststimulus responses that exceeded two SDs of the prestimulus baseline mean of the event-related average. The algorithm measured reflex latency (s), duration (s), magnitude (cents), and peak time (s). Response latency was the time of F_0 departure from the baseline mean by more than 2 SDs, while reflex duration reflected the total time a response remained above 2 SDs. Reflex magnitude (cents) was the maximum F_0 fluctuation from the mean prestimulus baseline F_0 beyond 2 SDs, and response peak time (s) reflected the poststimulus time corresponding to response magnitude. A minimum duration of at least 50 ms, a minimum peak time of 120 ms, a minimum magnitude of 5 cents, and a maximum latency no greater than 400 ms were required for a response to be considered valid. Previous experiments utilized similar duration criteria (Burnett *et al.*, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000; Larson *et al.*, 2001, 2000) to reduce invalid responses. Peak time, magnitude, and latency criteria were used to eliminate extreme data outliers. If no valid responses were observed for a given event-related average, then dependent measures were assigned a value of 0 and classified as a “non-response.” If more than one valid response occurred per event-related average, the response with the shortest latency was recorded.

While “non-responses” would be discarded from statistical analyses in most studies, we included them in our analyses so as to obtain a more accurate measure of overall variability. Given the relatively small number of these “non-responses,” their inclusion in our analysis was not expected to alter averaged data. Nevertheless we felt it was important to maintain the spirit of the data distribution given that the study was designed to test methodological adjustments. Subsequently, each “non-response” was transformed to the cell mean to avoid problems due to missing data in the statistical analyses of the dependent measures.

Logarithmic transformations were performed on the adjusted dependent measures to meet assumptions for parametric statistical comparison in repeated measures (within-subject) designs (i.e., equal samples per cell, minimal outliers, normal distribution, homogeneity of variance, compound symmetry, and sphericity). Transformed latency, peak time, and magnitude measurements were tested using repeated-measures ANOVAs (RM ANOVA) with a Bonferroni corrected alpha set at $p=0.01$. *Posthoc* Sheffé tests were used for follow-up analyses when needed. Response prevalence was tested with a nonparametric Cochran’s Q test across experimental conditions.

III. RESULTS

Figure 1 illustrates data for a representative subject across three of the four ISI conditions (2.5, 1.0, and 0.5 s). The four vertically stacked traces associated with each experimental condition represent the average responses obtained from 10, 15, 20, and 30 trials. Thin black lines represent average response and thick gray bars represent SE of the mean of all trials comprising the average. The vertical gray box represents the onset and duration of the pitch-shift stimuli. The overall morphology of responses is consistent across conditions and number of averaged trials. While the largest responses for this subject were observed for the 1.0-s ISI condition, we did not observe this trend across all subjects. On average the overall median reflex latency, peak time, and magnitude calculated across all subjects and collapsed across all conditions were 0.135 s, 0.219 s, and 14.34 cents, respectively.

The following RM ANOVA analyses were used to assess the response properties of the PSR across ISI conditions (2.5, 1.0, and 0.5) and averaging conditions (10, 15, 20, and 30). Results of the analysis of response latency did not reveal any statistically significant effects as a function of the ISI condition [$F(2,60)=2.45$, $p>0.05$] or the number of averaged responses [$F(3,90)=2.72$, $p>0.05$]. Likewise, peak-time analysis also did not indicate any statistically significant effects across ISI conditions [$F(2,60)=4.88$, $p>0.01$] or number of averages [$F(3,90)=0.85$, $p>0.05$]. Thus, temporal measures of latency and peak time were not affected by ISI rates as fast as 0.5 s or averages consisting of ten trials [Figs. 2(a) and (b)]. However, analysis of response magnitude revealed a significant main effect across the number of average responses [$F(3,90)=11.30$, $p<0.0001$], but not across ISI conditions [$F(2,60)=3.9$, $p>0.01$]. *Posthoc* Sheffé testing revealed larger magnitude responses for averages of 10 stimuli compared to averages of 15, 20, or 30 stimuli. Thus, larger magnitude responses were the result of averaging too few trials [Fig. 2(c)]. Interaction effects were not statistically significant.

In order to assess the effects of presentation of multiple pitch-shift stimuli compared to single stimulus presentation, data were also compared across the group of ten subjects exposed to four ISI conditions (5.0, 2.5, 1.0, and 0.5). Figure 3 displays data from a representative subject. Overall, there were no observable differences in reflex latency as a function of ISI [$F(3,27)=0.60$, $p>0.01$] or averaging condition [$F(3,27)=0.62$, $p>0.05$] as tested with a RM ANOVA ($N=10$). Likewise, no main effects or interactions were observed for measures of peak time across ISI [$F(3,27)=0.54$, $p>0.05$] or averaging conditions [$F(3,27)=0.73$, $p>0.05$]. In comparison, an increase in response magnitude occurred with an apparent decrease in the number of averages [$F(3,27)=14.14$, $p<0.0001$], but there were no statistically significant differences in magnitude as a function of ISI condition [$F(3,27)=0.31$, $p>0.05$]. *Posthoc* Sheffé analyses revealed that averages of 10 trials were greater in magnitude than averages of 15, 20, or 30 trials. Thus, PSR magnitude appears to be inversely related to the number of trials comprising the average. However, increasing the rate of stimulation from one per vocalization up to ten per vocalization (ISI 0.5 s) did not impact the pitch-shift reflex parameters.

Across all subjects there were 412 possible averaged responses [(21 subjects $\times$ 3 ISI $\times$ 4 averages) + (10 subjects $\times$ 4 ISI $\times$ 4 averages)]. Overall, 341 responses (83%) decreased in F_0 (opposing responses), 17 responses (4%) increased in F_0 ("following" responses), and 54 (13%) were considered "non-responses." The percentage of "following" responses and "non-responses" appeared to increase as a result of the decrease in interstimulus interval (Table I) with the highest incidence observed in the 0.5-s ISI condition. Similarly, an increase in the percentage of "non-responses" was observed as a result of a decrease in the number of averages (Table II) with the highest incidence observed for averages consisting of only 10 trials. Table III displays response prevalence as a percentage of the total responses across number of averages and ISI conditions. In general, fewer responses were observed for the 0.5-s ISI condition regardless of the number of responses comprising the average. Similarly, averaging of 10 responses

produced a lower percentage of responses than averages of 15, 20, or 30 trials across each of the ISI conditions. However, nonparametric statistical analysis of these response prevalence trends did not reach statistical significance (Cochran's $Q=11.22$, $p>0.05$) given the relatively small subject size.

IV. DISCUSSION

The primary objective of this study was to determine if the pitch-shift reflex could be elicited repetitively at relatively short ISIs during the course of a single vocalization. The secondary objective was to determine the number of trials needed to yield reliable averaged responses. A byproduct of both objectives was to determine if studies of the pitch-shift reflex could be conducted such that the time required for testing each subject could be reduced and thereby minimize potential fatiguing factors. It was reasoned that reduction of vocal fatigue would produce less variability in voice F_0 and hence more reliable data.

Determination of the optimal ISI for the study of event-related potentials is important for two reasons. The first is that studying the effects of ISI on response measures can provide indirect evidence of response processing duration. The second is that knowing the optimal ISI allows experiments to be conducted more rapidly than if an ISI is too long. Interstimulus intervals that are longer than the response processing time do not affect response measures, but increase the amount of experimentation time. Interstimulus intervals that are too short in duration may cause a reduction in response magnitude (Nelson and Lassman, 1968; Wikström *et al.*, 1996) regardless of the faster experimentation time. In the present study, reduction of the ISI from 5.0 to 0.5 s had no statistically significant effect on latency, peak time, or magnitude of the response, and thus we believe that the central processing mechanisms of this reflex are less than 0.5 s in duration. However, we could not assess the nature of the pitch-shift reflex with an ISI of less than 0.5 s because of limitations inherent in our averaging technique. Specifically, a prestimulus baseline period of 0.2 s is needed to obtain a stable baseline F_0 (approximately 20–50 vocal cycles). Further, a 0.5-s poststimulus window is needed to verify that the pitch-shift response returns to the baseline F_0 level. Nevertheless, our findings indicate the audio-vocal system adequately compensates for pitch fluctuations every 0.5 s during a sustained vocalization, which corresponds to corrective control at a rate of at least 2 Hz.

Even though the response measures were not affected by the reduction in the ISI, the shortest ISI (0.5 s) did elicit a greater percentage of “non-responses” and “following” responses (Tables I and II). We speculate that the decrease in response prevalence may have been caused by an overlap between the refractory period of one response and the prestimulus baseline of the next response. To illustrate this point for one set of data from a representative subject (Fig. 4), the poststimulus averaging window was extended to 1.0 s to encompass two sequential responses. Here it can be seen that the termination of the first reflex occurred at the same time as the prestimulus baseline period of the next reflex. The decreased magnitude of the second response suggests there was not a sufficient recovery period between successive stimuli presented with an ISI of 0.5 s. Thus, the overlap between the refractory period following one response and the prestimulus baseline of the next response may have lead to an increase in overall variability and a subsequent decrease in response prevalence in the average waveform. These data suggest that the ISI should be longer than 0.5 s so as to minimize potential response degradation.

The other goal of the present study was to determine the optimal number of trials needed to yield reliable and consistent averaged responses. Statistical analysis showed no difference in response latency or peak time regardless of the number of trials included in the averaged response. However, we observed larger magnitude responses and greater instances of “non-responses” when only 10 responses were averaged compared to 15, 20, or 30. It is difficult to explain why responses should have been larger with an average of only ten trials. However,

one possibility is that with relatively few responses comprising the average waveform, there was larger variability in the background noise level that added to the average response waveform yielding a larger response magnitude [Fig. 2(c)]. With a greater number of averages, a greater amount of noise cancellation occurs, and the measured response magnitude may be a more accurate estimate of the actual response as it regresses to the mean. Pursuing this logic, the greater incidence of “non-responses” with ten averages is most likely due to the fact that the “noise” in the background signal was not adequately averaged out. In other words, the response was no larger than the noise (variability) in the averaged signal. For these reasons, we conclude that at least 15 responses should be averaged per condition. Although our results did not show increased improvement with averages of 20 or 30 responses, it may be suggested that in some cases of extreme noise a greater number of averages would be warranted to reduce variability.

Previous studies in our lab using the pitch-shift paradigm have presented stimuli at an ISI of 5.0 s (Burnett *et al.*, 1998; Hain *et al.*, 2001; Kiran and Larson, 2001; Larson *et al.*, 2000) or once per vocalization. Subjects were instructed to vocalize at least 15–20 times per experimental condition to obtain an averaged response comprised of 15–20 trials. The response measures (latency, peak time, and magnitude) from the present study are very similar to those previously obtained. Therefore, it is suggested that similar data could have been obtained using only three vocalizations (five stimuli per vocalization). Although the median response magnitude (~14 cents) measured in the present study was less than values reported in the above studies, response magnitude is in part dependent on stimulus duration. The duration in this study (0.1 s) was substantially less than in most of the above studies, and thus leads to smaller magnitude responses than those previously reported. It must be emphasized, however, that if longer stimulus durations are used, the ISI would have to be increased to accommodate a longer response duration, and this could in turn reduce the number of stimuli per each vocalization, with a further increase in the number of vocalizations required per condition. Although short stimulus durations with short ISIs and repetitive trials per vocalization are appropriate for investigating reflexive properties of the audio-vocal system (Burnett *et al.*, 1998; Kiran and Larson, 2001; Larson *et al.*, 2000), longer durations with a single stimulus per vocalization are still preferable for other studies interested in more voluntary mechanisms of audio-vocal control (Hain *et al.*, 2001, 2000), or possibly studies of speech and prosody (Donath *et al.*, 2002; Natke and Kalveram, 2001).

V. CONCLUSION

Measuring physiologic responses to unanticipated perturbations in sensory feedback provides important information about neural control mechanisms for a particular motor task. Inferences regarding the neural mechanisms of motor planning, initiation, and corrective control can be made based on the timing, magnitude, and form of responses to perturbed sensory feedback in various tasks. Using this technique for the study of voice, it has been shown that auditory feedback is used to help stabilize voice F_0 around an intended level. In order to more fully appreciate the role of auditory feedback on voice control, a detailed understanding of interactions between stimulus variables and vocal responses is necessary. The present study has shown that the audio-vocal system can respond to repetitive pitch perturbations during a sustained vocalization. Furthermore, presenting pitch-shift stimuli with an ISI of at least 1.0 s and averaging at least 15 trials yields measures of response latency, peak time, magnitude, and prevalence comparable to those obtained at longer ISI and with greater number of averaged stimuli. Shorter ISIs and lower number of averages elicit more “non-responses” due to potential increased variability and overlap of successive averaging windows. Thus a five-fold reduction in subject participation time compared to previous studies in our lab may be obtained by reducing the ISI from 5.0 to 1.0 s. This reduction has the benefit of decreasing the total number

of vocalizations, and thus reduces confounding effects from subject fatigue, boredom, lapse of attention, and voice F_0 variability across vocalizations.

Acknowledgements

Research partially supported by NIH Grant No. DC02764-01 awarded to Dr. Charles R. Larson, Northwestern University, Evanston, IL. We would like to gratefully acknowledge the assistance of Ciara Leydon for suggestions regarding manuscript preparation, as well as invaluable feedback provided by two reviewers on an earlier version of the manuscript.

References

- Abbas PJ, Brown CJ. Electrically evoked auditory brainstem response: Refractory properties and strength-duration functions. *Hear Res* 1991;51:139–147. [PubMed: 2013540]
- Abbs JH, Gracco VL. Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *J Neurophysiol* 1984;51:705–723. [PubMed: 6716120]
- Barlow SM, Bradford PT. Comparison of perioral reflex modulation in the upper and lower lip. *J Speech Hear Res* 1996;39:55–75. [PubMed: 8820699]
- Burnett TA, Larson CR. Early pitch shift response is active in both steady and dynamic voice pitch control. *J Acoust Soc Am* 2002;112:1058–1063. [PubMed: 12243154]
- Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F_0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 1998;103:3153–3161. [PubMed: 9637026]
- Dawson GD. A summation technique for the detection of small evoked potentials. *Electroencephalogr Clin Neurophysiol* 1954;6:65–84. [PubMed: 13141922]
- Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables. *J Acoust Soc Am* 2002;111:357–366. [PubMed: 11831808]
- Gracco, VL. Central and peripheral components in the control of speech movements. In: Bell-Berti, F., editor. *Producing Speech: Contemporary Issues*. American Institute of Physics; Woodbury, NY: 1995. p. 417–431.
- Hain TC, Burnett TA, Larson CR, Kiran S. Effects of delayed auditory feedback (DAF) on the pitch-shift reflex. *J Acoust Soc Am* 2001;109:2146–2152. [PubMed: 11386566]
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res* 2000;130:133–141. [PubMed: 10672466]
- Jürgens U. A study of the central control of vocalization using the squirrel monkey. *Med Eng Phys* 2002;24:473–477. [PubMed: 12237042]
- Kiran S, Larson CR. Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's disease patients and normal controls. *J Speech Lang Hear Res* 2001;44:975–987. [PubMed: 11708537]
- Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparisons of voice F_0 responses to pitch-shift onset and offset conditions. *J Acoust Soc Am* 2001;110:2845–2848. [PubMed: 11785786]
- Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift onset velocity on voice F_0 responses. *J Acoust Soc Am* 2000;107:559–564. [PubMed: 10641664]
- Luthe L, Hausler U, Jürgens U. Neuronal activity in the medulla oblongata during vocalization. A single-unit recording study in the squirrel monkey. *Behav Brain Res* 2000;116:197–210. [PubMed: 11080551]
- Nakamura M, Nishida S, Shibasaki H. Evaluation of the signal-to-noise ratio for average evoked potentials: Determination of interstimulus interval and averaging number. *Front Med Biol Eng* 1989;1:341–349. [PubMed: 2486921]
- Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J Speech Lang Hear Res* 2001;44:577–584. [PubMed: 11407562]
- Nelson DA, Lassman FM. Effects of intersignal interval on the human auditory evoked response. *J Acoust Soc Am* 1968;44:1529–1532. [PubMed: 5702027]
- Saltzman E, Löfqvist A, Kay B, Kinsella-Shaw J, Rubin P. Dynamics of intergestural timing: A perturbation study of lip-larynx coordination. *Exp Brain Res* 1998;123:412–424. [PubMed: 9870601]

- Sapir, S.; McClean, MD. Evidence for acoustico-laryngeal and trigemino-laryngeal reflexes in man. American Speech-Language-Hearing Association; Los Angeles, CA: 1981.
- Shaiman S, Gracco VL. Task-specific sensorimotor interactions in speech production. *Exp Brain Res* 2002;146:411–418. [PubMed: 12355269]
- Smith A, McFarland DH, Weber CM, Moore CA. Spatial organization of human perioral reflexes. *Exp Neurol* 1987;98:233–248. [PubMed: 3666076]
- Solls MM, Brainard MS, Hessler NA, Doupe AJ. Song selectivity and sensorimotor signals in vocal learning and production. *Proc Natl Acad Sci USA* 2000;97:11836–11842. [PubMed: 11050217]
- Suthers RA, Goller F, Wild JM. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proc Natl Acad Sci USA* 2002;99:5680–5685. [PubMed: 11943843]
- Wikström H, Huttunen J, Korvenoja A, Virtanen J, Salonen O, Aronen H, Ilmoniemi RJ. Effects of interstimulus interval on somatosensory evoked magnetic fields (SEFs): A hypothesis concerning SEF generation at the primary sensorimotor cortex. *Electroencephalogr Clin Neurophysiol* 1996;100:479–487. [PubMed: 8980411]

FIG. 1.

Event-related averages of voice F_0 for a representative subject tested with ISI of 2.5, 1.0, and 0.5 s. Under each rate, averages of 10, 15, 20, and 30 responses were calculated (vertically stacked traces). Black lines represent the overall average response. Gray bars represent standard error of the mean for all responses comprising an average on a point-by-point basis. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0 s. Vertical line scales=20 cents.

FIG. 2.

(a) Median box-plots of response latency (s) by ISI and number of averages ($N=31$). (b) Median box-plots of responses peak time (s). (c) Median box-plots of response magnitude (cents). Box-plot definitions: The horizontal line through a box is the median. The shaded region surrounding the median is the 95% confidence interval. The upper and lower limits of the box represent the 75th and 25th percentiles, respectively. Whiskers extend to upper and lower limits of the main body of data. Points depicted by a circle are considered to be extreme data values, while very extreme values are plotted as asterisks.

FIG. 3.

Event-related averages of voice F_0 for a representative subject tested with ISI of 5.0, 2.5, 1.0, and 0.5 s. Under each rate, averages of 10, 15, 20, and 30 responses were calculated (vertically stacked traces). Black lines represent the overall average response. Gray bars represent standard error of the mean for all responses comprising an average on a point-by-point basis. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0 s. Vertical line scales=20 cents.

FIG. 4.

Event-related average for a representative subject displaying an overlap of successive pitch-shift reflex responses due to the short ISI at 0.5 s. Black lines represent the overall average voice F_0 and arrows denote response onset. Duration of the post-stimulus averaging window was extended to 1.0 s. Shaded boxes represent pitch-shift stimulus onset and duration (0.1 s) beginning at time 0.0 and 0.6 s.

TABLE I

Percentage (%) of “following” (FOLL), opposing (OPP), and “non-responses” (NR) across ISI condition.

ISI condition	5.0	2.5	1.0	0.5	Total
FOLL	2	2	3	8	4
OPP	90	88	86	73	83
NR	8	10	11	19	13

TABLE II

Percentage (%) of “following” (FOLL), opposing (OPP), and “non-responses” (NR) across number of averages.

No. of averages	10	15	20	30	Total
FOLL	3	3	6	5	4
OPP	80	84	82	84	83
NR	17	13	12	11	13

TABLE III

Response prevalence displayed as a percentage (%) of valid responses per cell (total number of opposing plus following responses divided by the total number of possible responses) across number of averages and ISI conditions. Weighted row and column means are based on raw data.

ISI	No. of averages				Weighted row mean
	10% Response	15% Response	20% Response	30% Response	
1.0	90	90	100	90	93
2.5	81	94	90	94	90
1.0	84	87	90	94	89
0.5	81	81	81	81	81
Weighted column mean	84	87	88	89	