Published in final edited form as:

J Acoust Soc Am. 2005 February ; 117(2): 850–857.

Voice responses to changes in pitch of voice or tone auditory feedback^{a)}

Mahalakshmi Sivasankar,

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

Jay J. Bauer,

Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413

Tara Babu, and

5403 MacArthur Boulevard, Washington, D.C. 20016

Charles R. Larson^{b)}

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

Abstract

The present study was undertaken to examine if a subject's voice F_0 responded not only to perturbations in pitch of voice feedback but also to changes in pitch of a side tone presented congruent with voice feedback. Small magnitude brief duration perturbations in pitch of voice or tone auditory feedback were randomly introduced during sustained vowel phonations. Results demonstrated a higher rate and larger magnitude of voice F_0 responses to changes in pitch of the voice compared with a triangular-shaped tone (experiment 1) or a pure tone (experiment 2). However, response latencies did not differ across voice or tone conditions. Data suggest that subjects responded to the change in F_0 rather than harmonic frequencies of auditory feedback because voice F_0 response prevalence, magnitude, or latency did not statistically differ across triangular-shaped tone or pure-tone feedback. Results indicate the audio–vocal system is sensitive to the change in pitch of a variety of sounds, which may represent a flexible system capable of adapting to changes in the subject's voice. However, lower prevalence and smaller responses to tone pitch-shifted signals suggest that the audio–vocal system may resist changes to the pitch of other environmental sounds when voice feedback is present.

I. INTRODUCTION

Despite the importance of vocalization for speech and singing, and the prevalence of disorders affecting the voice, neural mechanisms controlling the voice are poorly understood. As with other types of motor behaviors, researchers have speculated on the role of central neural mechanisms including sensory feedback for voice control. Data from studies using nerve stimulation and anesthetization techniques suggest that kinesthetic receptors are important in fine control of voice fundamental frequency (F_0) (Jürgens and Kirzinger, 1985; Ludlow *et al.*, 1992; Sundberg *et al.*, 1993). However, there is considerably more evidence for audition

^{a)}Material originally presented in “Comparison of vocal responses to changes in pitch of voice, triangular, and sinusoidal tone feedback,” The 3rd International Conference for Vocal Fold Physiology and Biomechanics, Denver, Colorado, September 2002. Portions of this manuscript were also presented as “Voice responses to changes in pitch of voice or tone feedback,” at the Speech Motor Control Conference, Williamsburg, VA, March, 2002.

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

as an important form of sensory feedback in vocal control, as seen by increased variability in voice F_0 and intensity following post-lingual deafness (Cowie and Douglas-Cowie, 1992), when speaking in noisy environments (Lane and Tranel, 1971), and under conditions where auditory feedback is delayed (Lechner, 1979) or masked with noise (Elliott and Niemoeller, 1970; Mürbe *et al.*, 2002; Ternström *et al.*, 1988).

The noninvasive pitch-shifting technique is a powerful method of assessing the role of auditory feedback in real-time vocal control. Using this technique, researchers have demonstrated that auditory feedback helps in the stabilization of voice F_0 through a closed-loop negative feedback mechanism with latencies between 100–150 ms (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2001). When the perceived voice pitch is greater than the intended pitch, voice F_0 is reduced to compensate for the disparity. Conversely, when the feedback pitch is perceived lower, voice F_0 is increased. A compensatory pitch-shift response has been demonstrated following pitch perturbations during sustained vocalizations (Burnett *et al.*, 1998; Hain *et al.*, 2000; Jones and Munhall, 2002; Kawahara, 1995; Larson *et al.*, 2000), whistling (Anstis and Cavanagh, 1979), glissandos (Burnett and Larson, 2002), and nonsense syllables (Donath *et al.*, 2002; Natke *et al.*, 2003; Natke and Kalveram, 2001). The pitch-shifting technique is a useful method to manipulate auditory feedback in real time, and provides a means to directly investigate the relationship between auditory feedback and voice F_0 control during ongoing vocalizations.

Changes in voice F_0 have also been observed in response to nonvocal sounds such as claps (Baer, 1979) and clicks (Sapir *et al.*, 1983b), suggesting that the audio–vocal system may be sensitive to other sounds besides the subject’s voice. However, these studies had limitations in that the stimuli were short in duration and were presented suddenly. Nevertheless, the question remains as to which specific acoustical properties of the stimuli elicited the vocal responses. Moreover, the response latencies to nonvocal sounds (50–60 ms) were shorter than those observed with the pitch-shift technique for voice auditory feedback perturbations (100–200 ms) (Burnett *et al.*, 1998; Donath *et al.*, 2002; Hain *et al.*, 2000; Larson *et al.*, 2000; Natke and Kalveram, 2001), suggesting that the mechanisms underlying the auditory–laryngeal reflex may be different than those underlying the pitch-shift response.

Exploring the potential influence of nonvocal sounds on voice F_0 regulation is important for at least two reasons. First, considering that auditory feedback has a regulatory function in voice control (Burnett *et al.*, 1998; Hain *et al.*, 2000), knowing whether certain ranges of F_0 or specific harmonic frequencies of the feedback signal are involved in this regulatory process would enable us to understand how the auditory system interacts with laryngeal and respiratory motor systems for vocal control. Second, understanding the influence of nonvocal sounds on voice control is important for learning how the voice is controlled in a noisy environment while singing or speaking. Depending on the task, singers either maintain their intended vocal pitch by ignoring auditory feedback from other vocalists and instruments, or by matching their pitch to auditory feedback.

In the present study we compared pitch-shift responses elicited by perturbations in pitch of either voice feedback or nonvoice sounds to ascertain whether pitch-shift responses are sensitive only to the voice or to other periodic sounds as well. A secondary question was whether the audio–vocal system was sensitive primarily to the F_0 of the signal or a combination of the F_0 and harmonic components. To address these questions we compared pitch-shift responses elicited by perturbed voice feedback, perturbed triangular-shaped waves, and perturbed sinusoidal waves across two experiments. We hypothesized that the audio–vocal system would be more sensitive to the subject’s own voice feedback compared to the tone feedback, as reflected in greater response prevalence and magnitudes for voice compared to the tone signals. Moreover, we predicted that the triangular signal, with its many harmonics

and thus greater similarity to the voice, would elicit greater response prevalence and magnitudes than the sinusoidal signal.

II. METHODS

Experiment 1 compared responses to pitch-shifted voice and pitch-shifted complex tone (triangle) auditory feedback. Experiment 2 compared responses to pitch-shifted voice and pitch-shifted pure-tone (sinusoidal) auditory feedback.

A. Experiment 1

1. Subjects—Nineteen healthy female subjects (21–36 years of age; mean age = 25.5 years) participated in this study. Previous studies using perturbed voice auditory feedback have revealed no differences in voice F_0 response measures between male and female subjects (unpublished observations). In order to maintain a similar relationship between voice F_0 and harmonics of the voice and tone feedback across all subjects, only female subjects were tested in this experiment. All subjects passed a hearing screening at 15-dB SPL (octave frequencies between 500–8000 Hz), and reported no history of neurological deficits, speech, language, or voice disorders. None of the subjects were trained singers.

2. Apparatus and procedures—Subjects were seated comfortably in a sound-treated booth and instructed to vocalize /u/ vowel sounds at a comfortable and steady habitual pitch while listening in near-real time to pitch-shifted voice and/or tone auditory feedback over headphones. Subjects were instructed to maintain their voice pitch regardless of any auditory feedback disturbance they might hear. Subjects sustained each vocalization for approximately 5 seconds at 77 dB SPL (self-monitored visually with a Dorrrough loudness monitor model 40-A). Each subject produced 48 vocalizations across four blocks of 12 consecutive vocalizations. Vocalizations were transduced with an AKG boom-set microphone (HSC 200; microphone-to-mouth distance of 5 cm), and then amplified by a Mackie mixer (model 1202). A 534-Hz triangular tone (TT) produced by a function generator (Wavetek model 188) served as the external auditory tone. A 534-Hz tone was selected as the nonvoice sound after pilot study data. These data revealed very little overlap between the fundamental and harmonic frequencies of the 534-Hz tone with the harmonics of a female voice produced at habitual pitch. Voice and tone signals were processed for pitch shifting using an Eventide Ultraharmonizer (SE 3000) controlled by Musical Instrument Digital Interface software (MAX v3.5.9 by Opcode). During each vocalization beginning 500–1000 ms after vocal onset, the Ultraharmonizer was triggered five times in succession to increase or decrease voice or tone feedback pitch by 100 cents (equal to one semitone) each for a duration of 300 ms with a minimum interstimulus interval of at least 600 ms between perturbations. The onset of each pitch-shift stimulus was rapid, with a rise time of approximately 10 ms. To partially mask bone-conducted feedback and reduce potential binaural beating with tone feedback, voice and tone signals were mixed (Mackie mixer model 1202-VLZ), with 70-dB SPL pink noise (Goldline Audio Noise Source, model PN2; spectral frequencies 1 to 5000 Hz). Feedback signals were amplified with a Crown audio amplifier (D75-A), routed through HP decibel attenuators (model 350D), and fed back to the subjects via circumaural AKG headphones (model HSC 200). Voice and tone loudness feedback, when amplified, were perceptually equalized at about 88 dB SPL. Acoustical equipment was calibrated with a Brüel & Kjær 2203 sound-level meter (weighting A) prior to data collection. The “A”-weighting scale was used because of low-level hum of the ventilation system and rack-mounted equipment in the building and laboratory immediately adjacent to the sound-attenuated booth.

Each subject participated in four experimental conditions in a randomized order. Subjects vocalized sustained /u/ vowels and heard

1. only their amplified voice feedback, which was subsequently pitch shifted (V_{ps}).
2. only amplified triangle tone feedback, which was subsequently pitch shifted (TT_{ps}). In this case, the subject's voice feedback was not amplified or pitch shifted, so the primary source of auditory feedback was the tone. However, subjects may have been able to perceive the voice signal through the bone-conducted pathway.
3. both their amplified voice feedback and triangle tone feedback, but only their voice feedback was pitch shifted ($V_{ps} + TT$).
4. both their amplified voice feedback and triangle tone feedback, but only the triangle tone feedback was pitch shifted ($V + TT_{ps}$).

B. Experiment 2

1. Subjects—Nineteen healthy female subjects (19–30 years of age; mean age = 23 years) who did not participate in experiment 1, were tested in this study. Subject selection criteria were identical to experiment 1.

2. Apparatus and procedures—The methodology was similar to experiment 1; however, the external auditory tone was a 534-Hz pure-tone (PT) stimulus (Wavetek function generator model 188). Four conditions were tested for each subject in a random order: V_{ps} , PT_{ps} , $V_{ps} + PT$, $V + PT_{ps}$.

3. Data analysis—In each experiment for each condition, the subject's vocal signal, auditory feedback, control pulse indicating direction of the pitch shift, and TTL trigger pulse indicating onset of the pitch-shift stimulus were recorded. Signals were digitized on-line onto a laboratory computer at a 12-bit sampling rate of 10 kHz (5-kHz antialias filtering) using A/D conversion software (MACLAB CHART v3.5 by AD Instruments).

In off-line analysis (IGOR PRO software, version 4.0 by Wavemetrics, Inc.) the voice signal was low-pass filtered at 200 Hz, differentiated, and then smoothed with a five-point binomial, sliding window so as to remove high-frequency harmonics from the audio signals. A wave representing temporal fluctuations in voice F_0 (where voltage corresponded to F_0 in hertz) was then extracted from these preprocessed signals using a custom software algorithm in IGOR PRO. This algorithm detected positive-going threshold-voltage crossings, interpolated the time fraction between the two sample points that constituted each crossing, and calculated the reciprocal of the period defined by the center points. These F_0 signals were then converted to a cents scale using the following equation: $\text{cents} = 100(12(\log_{10}(f_2/f_1))/\log_{10}(2))$, where f_1 equals an arbitrary reference note at 195.997 Hz (G4) and f_2 equals the voice signal in hertz. The conversion of voltage signals to cents permitted a relative comparison of the extent of change in the F_0 in response to a pitch modulation across all conditions, regardless of baseline F_0 (Baken and Orlikoff, 2000). Voice signals were then low-pass filtered at 20 Hz to remove sharp discontinuities associated with each glottal cycle. Event-related averages were generated for each subject and for each experimental condition by time-aligning the voice signals with pitch-shift stimulus onset (TTL control pulse). Event-related averages consisted of a minimum of 15 trials, each with a 400-ms prestimulus baseline and a 600-ms poststimulus response window. A pitch-shift response was automatically identified using a custom software algorithm, and the pitch-shift response magnitude and latency were concurrently measured from each event-related average. Pitch-shift responses were identified as valid or invalid responses according to the criteria below. A valid response was defined as a deviation from the averaged F_0 baseline trace with a magnitude ≥ 2 standard deviations (SDs) of the prestimulus baseline, a peak time ≥ 120 ms and a latency ≥ 50 ms but ≤ 400 ms after stimulus onset.

4. Statistical analysis—The total number of trials in each experiment equaled 152 (19 subjects $\times$ 4 feedback conditions $\times$ 2 pitch-shift stimulus directions). To determine if the total number of valid responses (response prevalence) differed across conditions, nonparametric Cochran's Q was calculated separately for each experiment across the four conditions collapsed across stimulus direction. Previous studies have indicated that responses to downward pitch shifts do not differ quantitatively from responses elicited by upward pitch-shifts (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2001). Statistical analyses for pitch-shift response magnitude and latency were calculated only for valid responses in each experiment. Raw data for response magnitude and latency did not fulfill assumptions of normality (non-normal distribution, small sample size, and unequal samples per cell) and therefore Bonferroni corrected nonparametric repeated Friedman's ANOVAs were calculated for each variable (Siegel and Castellan, 1988). Additionally, all valid responses were categorized according to whether they changed in the opposite direction (compensatory response) or the same direction as the pitch-shift stimulus ("following" response) (Burnett *et al.*, 1998). The proportion of compensatory and following responses was then compared across all conditions in each experiment and submitted to significance testing using a chi-square analysis. Furthermore, response prevalence for pitch-shifted triangular (experiment 1) and pure tone (experiment 2) feedback were compared using a chi-square analysis. Response magnitude and latency were compared across experiments with Wilcoxon Mann–Whitney U tests.

III. RESULTS

A. Experiment 1

1. Response prevalence—Response prevalence differed significantly across the four conditions (Fig. 1). More valid pitch-shift responses were elicited for voice feedback pitch shifts (V_{ps}) than for perturbed triangular tone feedback (TT_{ps}) (Cochran's $Q = 12.66$, $df=3$, $p<0.05$). Response prevalence for voice feedback pitch shifts (V_{ps}), and voice feedback pitch shifts in the presence of tone feedback ($V_{ps} + TT$) were significantly greater than that for tone feedback pitch shifts in the presence of voice ($V + TT_{ps}$) (McNemar's *post hoc* tests: $\chi^2 = 5.56$, $df=1$, $p<0.05$; $\chi^2 = 4.96$, $df=1$, $p<0.05$, respectively).

2. Response magnitude and latency—Response magnitude differed significantly across voice and tone feedback pitch-shift conditions (Friedman's $F = 7.11$, $k = 3$, $p<0.05$, Fig. 2). *Post hoc* testing indicated response magnitude for voice feedback pitch shifts (23 ± 13 cents) and voice feedback pitch shifts in the presence of tone (20 ± 11 cents) were significantly greater than that for tone feedback pitch shifts (12 ± 6 cents; $p<0.05$). The response magnitudes for tone feedback pitch shifts in the presence of voice (10 ± 4 cents) were lower than that for the other three conditions, but were excluded from statistical analysis due to poor response prevalence (small proportion of valid responses). Response latency did not differ across experimental conditions (Friedman's $F = 1.34$, $k = 3$, $p>0.05$, Fig. 3). Mean response latency varied from 106 ms (voice feedback pitch shifts) to 113 ms (tone feedback pitch shifts). Mean response latency was lowest for tone feedback pitch shifts in the presence of voice (95 ms), but again this condition was excluded from analysis due to low response prevalence.

3. Response type—The proportion of compensatory and following responses did not vary across conditions as seen in Fig. 4 ($\chi^2 = 1.11$, $df = 1$, $p>0.05$). Across all four conditions, the majority of valid pitch-shift responses (81%) was compensatory. That is, the responses were in the opposite direction of the pitch-shift stimulus. A few responses (19%) were "following" or in the same direction as the pitch-shift stimulus. Although statistically nonsignificant, a greater proportion of following responses was observed for tone feedback pitch shifts in the presence of voice (29%) than for any other condition (less than 20%).

B. Experiment 2

1. Response prevalence—Results from experiment 2 using a pure tone as the external auditory stimulus were similar to experiment 1. The number of valid responses differed significantly across all four conditions (Cochran's $Q = 11.76$, $df = 3$, $p < 0.05$, Fig. 1), indicating that the pitch-shift response was not equally prevalent for voice and pure-tone feedback pitch shifts. Response prevalence for voice feedback pitch shifts and voice feedback pitch shifts in the presence of pure-tone feedback were significantly greater than that for pure-tone feedback pitch-shifts in the presence of voice (McNemar's *post hoc*: $\chi^2 = 6.32$, $df = 1$, $p < 0.01$ and $\chi^2 = 4.99$, $df = 1$, $p < 0.05$, respectively).

2. Response magnitude and latency—Response magnitude differed significantly across conditions as seen in Fig. 2 (Friedman's $F = 6.73$, $k = 3$, $p < 0.05$), indicating that the extent of pitch-shift response magnitude varied for voice and pure-tone pitch-shifts. Response magnitude for voice feedback pitch shifts (22 ± 13 cents) and voice feedback pitch shifts in the presence of pure tone (19 ± 13 cents) were significantly greater than that for pure-tone feedback pitch shifts (12 ± 6 cents; $p < 0.05$). The response magnitude for pure-tone feedback pitch shifts in the presence of voice (9 ± 4 cents) was lower than that for the other three conditions, but was excluded from statistical analysis due to the significantly greater proportion of invalid responses. As seen in Fig. 3, response latency did not differ across experimental conditions (Friedman's $F = 2.01$, $k = 3$, $p > 0.05$). Mean response latency varied from 110 ms (pure-tone feedback pitch shifts) to 121 ms (voice feedback pitch shifts). Mean response latency was lowest for pure-tone feedback pitch shifts in the presence of voice (105 ms).

3. Response type—The proportion of compensatory and following responses did not vary across conditions as seen in Fig. 4 ($\chi^2 = 1.21$, $df = 1$, $p > 0.05$). Across all four conditions, the majority of the pitch-shift responses (81%) was compensatory, but a few responses (19%) were following. Additionally, a greater proportion of following responses was observed for pure-tone feedback pitch shifts in the presence of voice (28%) than any other condition (less than 18%).

C. Experiment 1 and experiment 2

Pitch-shift response prevalence did not vary with the nature of external auditory tone (triangular or pure tone; see Table I). That is, between-experiment comparisons demonstrated comparable response prevalence for triangular and pure-tone feedback pitch shifts ($\chi^2 = 1.19$, $df = 1$, $p > 0.05$). Similarly, response prevalence did not differ for voice feedback pitch shifts in the presence of tone (triangular or sinusoidal $\chi^2 = 1.01$, $df = 1$, $p > 0.05$) or when tone (triangular or sinusoidal) was shifted in the presence of voice ($\chi^2 = 0.96$, $df = 1$, $p > 0.05$). The finding of reduced response prevalence for $V + T_{ps}$ condition across both experiments suggests that the audio-vocal system may rely more on voice feedback than on tone feedback for maintaining vocal F_0 stability. Response magnitude and response latency did not vary with triangular or pure tone (Wilcoxon Mann-Whitney statistic, $p > 0.05$), indicating that similar patterns of responses were elicited regardless of the type of external auditory tone. Furthermore, the proportion of compensatory and following responses was similar across both experiments, suggesting that the audio-vocal system responds in a similar fashion to complex and simple nonverbal external stimuli (Table I).

IV. DISCUSSION

The role of the pitch-shift response in stabilizing voice F_0 by correcting for pitch perturbations has been widely recognized (Burnett *et al.*, 1998; Burnett and Larson, 2002; Donath *et al.*, 2002; Hain *et al.*, 2000; 2001; Larson *et al.*, 2001; 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001). Additionally, there has been preliminary evidence for the presence of

reflexive changes in voice F_0 in response to short-duration nonverbal sounds such as claps (Baer, 1979) and clicks (Sapir *et al.*, 1983b). Others have reported that when the pitch of whistle is altered, subjects produce a compensatory response that is similar to the pitch-shift response (Anstis and Cavanagh, 1979). Compensatory voice F_0 responses have also been elicited by sinusoidal modulations of saw-tooth tones, suggesting the presence of a short latency brainstem auditory–laryngeal reflex pathway (Sapir *et al.*, 1983a). Whether the audio–vocal system responds to pitch changes in nonvoice tones in the presence of unperturbed voice feedback has not been examined before. The present study demonstrates a compensatory vocal response to pitch changes in external nonverbal auditory tones in the presence of amplified, unaltered voice feedback.

Regardless of the ability of the system to respond to both voice or nonvoice feedback, the audio–vocal system appears to depend more on voice feedback than tone feed-back to stabilize voice F_0 . This was demonstrated by greater response prevalence and magnitude for conditions where voice feedback was pitch shifted (V_{ps} , $V_{ps} + T$). Conversely, response prevalence was lowest in the condition where voice feedback was not pitch shifted ($V + T_{ps}$). That is, voice feedback but not tone feedback appears to be used as the guiding referent to correct for perceived changes in voice F_0 . Similarly, response magnitude was greatest for conditions where voice feedback was available and pitch shifted (V_{ps} , $V_{ps} + T$), and lowest for the condition where voice feedback was absent (T_{ps}) or unperturbed ($V + T_{ps}$). This increased sensitivity and dependence on voice would be necessary to prevent a person's voice from varying excessively with fluctuations in external environmental sounds. The lower response magnitude to pitch fluctuations in conditions where tone was shifted could reflect a mechanism employed by the audio–vocal system to avoid phonatory instability. If the audio–vocal system were equally sensitive to changes in voice and tone feedback, a person's voice would fluctuate with changes in the pitch of sounds in the surrounding environment. This selective sensitivity of the audio–vocal system to voice over tone may help explain how singers in a choir are able to maintain their vocal pitch in the presence of conflicting auditory feedback from adjacent singers or instruments. However, future studies are needed to address the role of unrelated voice auditory feedback on elicitation of voice F_0 responses. Nevertheless, our finding that subjects responded to pitch perturbations in the tone demonstrates that when a subject vocalizes in the presence of another acoustical signal, pitch perturbations in the external sound could destabilize voice F_0 . In other words, subjects would have greater difficulty holding a steady voice F_0 in the presence of external periodic sounds. Results from the present study suggest that the magnitude of the instability is small (e.g., 9–10 cents), but under some circumstances (perhaps extremely noisy conditions or underlying vocal pathology), the instability could be greater.

The difference in sensitivity to the voice and tone feedback signals may reflect developmental or environmental factors important for voice control. Humans are dependent on voice for speech, and the dominance of voice over tone may be an inherent feature of the audio–vocal system adapted to enhance communication effectiveness.

Although it is important to understand why responses to the two types of signals (voice and tone) differed, it is equally important to understand factors that led to similarities in responses. That is, subjects produced compensatory responses to both the voice and tone-shifted signals with similar response latency, even though the magnitude and prevalence to the tone-shifted signals were reduced. These similarities suggest that the audio–vocal system may respond to a common property of both simple and complex tones. The most likely candidate for such an acoustical property is F_0 . However, the 200–300-Hz difference in frequency between the tone (534 Hz) and the voice F_0 of the subjects would seem to be dissimilar enough, to obviate confusion of the tone for the subject's own voice feedback. All subjects verbally reported that the tone signal sounded much different than their voice. Moreover, results of pitch perception experiments indicate that people can detect differences in tone frequencies as small as 3 to 4

Hz (Wier *et al.*, 1977), suggesting that the 200- to 300-Hz difference between the voice and the tone signals in the present experiment should have been easy to discriminate. It is unlikely that overlap in harmonic frequencies of the tone and harmonics of the voice feedback accounted for the similarities in results reported here, because similar results were obtained with the pure-tone signal (no harmonic frequencies).

An alternative consideration is whether an overlap between harmonic frequencies of the voice signal and the F_0 of the tone signals accounted for the similarities in results. As a test of this possibility, we subjected 30% of our data to FFT analysis to determine if there was a pattern of overlap between the F_0 and harmonics, of the voice and the tones. By comparing the frequency components of voice output with that of the tone feedback, regions of overlap could be easily identified. The results of these FFT analyses showed that there was no systematic overlap between voice harmonics and tone frequencies for all 30% of data. Thus, overlap in frequency content (or masking of tone by voice) between the voice and tone signals does not appear to explain our data.

Since it seems unlikely that similarities in F_0 of the signals or overlaps of harmonic content of the signals can easily explain the similarity in responsiveness to the voice and tone signal, other explanations are needed. One possibility is that the audio-vocal system is a very adaptable system and will respond to changes in the voice as well as changes in frequency of any tone while a subject is vocalizing. Thus, the audio-vocal system may utilize a nonselective “pitch change detector” to respond to perturbations in pitch of periodic types of auditory feedback. Such adaptability would be advantageous in that it would allow the system to respond to the subject’s voice regardless of the F_0 . Through a process of development, the system could learn to respond to a subject’s voice at many different frequencies, and this could result in a system that is not selective to precise acoustical properties of feedback signals, only a change in frequency *per se*. One could also argue that since the control of whistling frequency is sensitive to perturbed pitch feedback (Anstis and Cavanagh, 1979), the system is capable of a motor response to perturbations in pitch of any frequency-modulating device that a person has learned to control. Such an outcome would predict that control of frequency in any musical instrument might also be sensitive to perturbations in pitch as well (Parlitz and Bangert, 1999). It is unlikely that startle responses could account for the results since upward and downward pitch-shift stimuli led to compensatory responses (71%–89% of the time) with an average latency between 95–124 ms. If the subjects were startled by the stimuli, one would expect short-latency, noncompensatory responses due to a contraction of laryngeal adductor and respiratory muscles resulting only in increases in subglottal pressure and voice F_0 . Moreover, startle responses habituate to repeated stimulation (Shalev *et al.*, 1992), while responses to pitch-shifted voice feedback do not.

Finally, in previous reports using the pitch-shifting technique, we have reported that some responses “follow” the stimulus direction while most oppose the perturbation direction (compensation). Such compensatory behavior and error correction have also been observed in other sensory and motor systems including the visual, auditory, and articulatory systems (Baum *et al.*, 1996; Cole and Abbs, 1988; Gracco and Abbs, 1985; Held, 1965; Shaiman and Gracco, 2002). In an investigation on the role of auditory adaptation in the speech domain, human subjects were provided auditory feedback in which the vowel formants being produced were shifted slowly over time (Houde and Jordan, 1998). Over 4220 trials, subjects adjusted vowel production to compensate for the vowel’s perturbed identity. In the present study we observed that most responses were compensatory and a small minority of them were of the following type. Moreover, there was a nonsignificant increase in numbers of following responses in conditions where the tone was shifted rather than the voice. It was previously noted that more following responses were observed with large magnitude stimuli (Burnett *et al.*, 1998), and it was speculated that feedback signals that differ from the voice could lead subjects to change

their choice of referent and follow the stimulus direction (Hain *et al.*, 2000). The larger numbers of following responses with tone stimuli in the present study support this speculation, but more data are needed before we can adequately explain such responses.

V. CONCLUSION

The present study tested the responses of the audio–vocal system to pitch perturbations in voice and external nonvoice sounds. Results demonstrate that the system responds with a robust compensatory response to pitch changes in both voice and external nonvoice sounds (triangular and pure tones). However, greater response magnitudes were observed to pitch-shifted voice feedback compared to pitch-shifted tone feedback, suggesting that the audio–vocal system depends chiefly on voice feedback to stabilize voice F_0 . Even in the presence of simultaneous voice and tone feedback signals, the system appears to rely more on voice feedback than external nonvoice feedback. Furthermore, the presence of a pitch-shift response to pitch perturbations in either a nonverbal complex or pure tone, suggests that the audio–vocal system is sensitive to changes in fundamental frequency of tonal feedback rather than the harmonic energy. The ability of the audio–vocal system to respond to both the subject's voice and external tone stimuli may reflect a flexible system capable of adapting to behavioral and developmental changes in vocal behavior.

Acknowledgements

This research was supported by NIH Grant No. DC006243-01A1.

References

- Anstis SM, Cavanagh P. "Adaptation to frequency-shifted auditory feedback,". *Percept Psychophys* 1979;26:449–458. [PubMed: 542360]
- Baer T. "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes,". *J Acoust Soc Am* 1979;65:1271–1275. [PubMed: 458049]
- Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular, San Diego).
- Baum SR, McFarland DH, Diab M. "Compensation to articulatory perturbation: Perceptual data,". *J Acoust Soc Am* 1996;99:3791–3794. [PubMed: 8655810]
- 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]
- Cole KJ, Abbs JH. "Grip force adjustments evoked by load force perturbations of a grasped object,". *J Neurophysiol* 1988;60:1513–1522. [PubMed: 3193168]
- Cowie, R., and Douglas-Cowie, E. (1992). "Postlingually acquired deafness," in *Trends in Linguistics, Studies and Monographs* (Mouton de Gruyter, New York).
- 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]
- Elliott L, Niemoeller A. "The role of hearing in controlling voice fundamental frequency,". *Int Aud* 1970;IX:47–52.
- Gracco VL, Abbs JH. "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes,". *J Neurophysiol* 1985;54:418–432. [PubMed: 4031995]
- 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]

- Held R. "Plasticity in sensory-motor systems,". *Sci Am* 1965;213(5):84–94. [PubMed: 5828465]
- Houde JF, Jordan MI. "Sensorimotor adaptation in speech production,". *Science* 1998;279:1213–1216. [PubMed: 9469813]
- Jones JA, Munhall KG. "The role of auditory feedback during phonation: Studies of mandarin tone production,". *J Phonetics* 2002;30:303–320.
- Jürgens U, Kirzinger A. "The laryngeal sensory pathway and its role in phonation. A brain lesioning study in the squirrel monkey,". *Exp Brain Res* 1985;59:118–124. [PubMed: 4018191]
- Kawahara, H. (1995). "Hearing voice: Transformed auditory feedback effects on voice pitch control," 'Computational Auditory Scene Analysis' and 'International Joint Conference on Artificial Intelligence,' Montreal.
- Lane H, Tranel B. "The Lombard sign and the role of hearing in speech,". *J Speech Hear Res* 1971;14:677–709.
- 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]
- 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]
- Lechner B. "The effects of delayed auditory feedback and masking on the fundamental frequency of stutterers and non-stutterers,". *J Speech Hear Res* 1979;22:243–253.
- Ludlow C, Van Pelt F, Koda J. "Characteristics of late responses to superior laryngeal nerve stimulation in humans,". *Ann Otol Rhinol Laryngol* 1992;101:127–134. [PubMed: 1739256]
- Mürbe D, Pabst F, Hofmann G, Sundberg J. "Significance of auditory and kinesthetic feedback to singers' pitch control,". *J Voice* 2002;16:44–51. [PubMed: 12002886]
- 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]
- Natke U, Donath TM, Kalveram KT. "Control of voice fundamental frequency in speaking versus singing,". *J Acoust Soc Am* 2003;113:1587–1593. [PubMed: 12656393]
- Parlitz D, Bangert M. "Short and medium motor responses to auditory pitch shift: Latency measurements of the professional musician's audio-motor loop for intonation,". *J Acoust Soc Am* 1999;105:1298.
- Sapir S, McClean M, Luschei ES. "Effects of frequency-modulated auditory tones on the voice fundamental frequency in humans,". *J Acoust Soc Am* 1983a;73:1070–1073. [PubMed: 6841812]
- Sapir S, McClean MD, Larson CR. "Human laryngeal responses to auditory stimulation,". *J Acoust Soc Am* 1983b;73:315–321. [PubMed: 6826901]
- Shaiman S, Gracco VL. "Task-specific sensorimotor interactions in speech production,". *Exp Brain Res* 2002;146:411–418. [PubMed: 12355269]
- Shalev AY, Orr SP, Peri T, Schreiber S, Pitman RK. "Physiologic responses to loud tones in Israeli patients with posttraumatic stress disorder,". *Arch Gen Psychiatry* 1992;49:870–875. [PubMed: 1444725]
- Siegel, S., and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).
- Sundberg, J., Iwarsson, J., and Billström, A.-M. H. (1993). "Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers," 22nd Annual Symposium Care of the Professional Voice, Philadelphia.
- Ternström S, Sundberg J, Colldén A. "Articulatory f_0 perturbations and auditory feedback,". *J Speech Hear Res* 1988;31:187–192. [PubMed: 3398492]
- Wier CC, Jesteadt W, Green DM. "Frequency discrimination as a function of frequency and sensation level,". *J Acoust Soc Am* 1977;61:178–184. [PubMed: 833369]

FIG. 1.

Bar graph displaying response prevalence (%) across four conditions for both experiment one (triangle tone) and two (pure tone). Filled bars represent mean % of response prevalence for subjects in experiment 1. Un-filled bars represent mean % of response prevalence for subjects in experiment 2. Abbreviations: V_{ps} = voice feedback with voice pitch shifted; T_{ps} = tone feedback with tone pitch shifted; $V_{ps} + T$ = voice and tone feedback with only voice pitch shifted; $V + T_{ps}$ = voice and tone feedback with only tone pitch shifted.

FIG. 2.

Bar graph displaying mean response magnitude (cents) across four conditions for experiment one (triangle tone) and two (pure tone). Filled bars represent mean response magnitude ± 1 SD across conditions for subjects in experiment 1. Unfilled bars represent mean response magnitude ± 1 SD across conditions for subjects in experiment 2. Abbreviations: same as Fig. 1.

FIG. 3.

Bar graph displaying mean response latency (ms) across four conditions for both experiment one (triangle tone) and two (pure tone). Filled bars represent mean response latency ± 1 SD across conditions for subjects in experiment 1. Unfilled bars represent mean response latency ± 1 SD across conditions for subjects in experiment 2. Abbreviations: same as Fig. 1.

FIG. 4.

Bar graph displaying response type (%) across four conditions for both experiment one (triangle tone) and two (pure tone). Filled bars represent % of compensatory responses in experiment 1. Filled horizontal striped bars represent % of following responses in experiment 1. Unfilled bars represent % of compensatory responses in experiment 2. Unfilled vertical striped bars represent % of following responses in experiment 2. Abbreviations: same as Fig. 1.

TABLE I

Comparisons of response characteristics: % of response prevalence, mean response magnitude (cents $\pm$ SD), mean response latency (ms $\pm$ SD), and % of response direction across the four conditions of V_{ps} (voice feedback with voice pitch shifted), T_{ps} (triangular or sinusoidal tone feedback with tone pitch-shifted), $V_{ps} + T$ (voice and tone feedback with only voice pitch shifted), and $V + T_{ps}$ (voice and tone feedback with only tone pitch shifted) in each of two experiments [triangle tone stimulus (TT), or pure-tone stimulus (PT)].

Conditions Experiment	V_{ps}		T_{ps}		$V_{ps} + T$		$V + T_{ps}$	
	TT	PT	TT	PT	TT	PT	TT	PT
Prevalence	82%	83%	74%	66%	82%	79%	34%	30%
Mean mag.	23	22	12	12	20	19	10	9
(± 1 SD)	(± 13)	(± 13)	(± 6)	(± 6)	(± 11)	(± 13)	(± 4)	(± 4)
Mean latency	105	124	113	110	108	110	95	105
(± 1 SD)	(± 39)	(± 45)	(± 45)	(± 40)	(± 33)	(± 45)	(± 35)	(± 45)
Compensatory Responses	80%	82%	89%	87%	81%	81%	71%	72%
Following Responses	20%	18%	11%	13%	19%	19%	29%	28%