Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex

Hanjun Liu and Charles R. Larson[a)]
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA

(Received 27 February 2007; revised 26 September 2007; accepted 28 September 2007)

The purpose of the present study was to investigate the responsiveness of the pitch-shift reflex to small magnitude stimuli and voice fundamental frequency (F_0) level. English speakers received pitch-shifted voice feedback (±10, 20, 30, 40, and 50 cents, 200 ms duration) during vowel phonations at a high and a low F_0 level. Mean pitch-shift response magnitude increased as a function of pitch-shift stimulus magnitude, but when expressed as a percent of stimulus magnitude, declined from 100% with ±10 cents to 37% with ±50 cents stimuli. Response magnitudes were larger and latencies were shorter with a high F_0 level (16 cents; 130 ms) compared to a low F_0 level (13 cents; 152 ms). Data from the present study demonstrate that vocal response magnitudes are equal to small perturbation magnitudes, and they are larger and faster with a high F_0 voice. These results suggest that the audio-vocal system is optimally suited for compensating for small pitch rather than larger perturbations. Data also suggest the sensitivity of the audio-vocal system to voice perturbation may vary with F_0 level.
[DOI: 10.1121/1.2800254]

PACS number(s): 43.70.Gr, 43.70.Fq, 43.70.Jt, 43.71.Bp [AL] Pages: 3671–3677

I. INTRODUCTION

Control of voice fundamental frequency (F_0) is an important issue in speech communication, and auditory feedback plays a significant role in this process. Numerous studies have demonstrated that subjects produce compensatory responses in voice F_0 following unanticipated perturbations in voice pitch feedback. The fact that the pitch-shift reflex functions to help stabilize voice F_0 have been demonstrated during sustained vowels (Bauer and Larson, 2003; Burnett *et al.*, 1998; Chen *et al.*, 2007; Hain *et al.*, 2000; Kawahara, 1995; Kiran and Larson, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005), glissandos (Burnett and Larson, 2002), speech (Bauer, 2004; Chen *et al.*, 2007; Xu *et al.*, 2004) and nonsense syllables (Donath *et al.*, 2002; Natke *et al.*, 2003; Natke and Kalveram, 2001).

Responses to pitch-shifted voice feedback are generally of two types. The far more prevalent type changes voice F_0 in the opposite direction to the pitch-shift stimulus and has been termed "opposing" response. It is thought that these responses correct for errors between intended F_0 and the actual F_0 produced. The second type, termed "following" response, is relatively rare and changes F_0 in the same direction as the stimulus. These responses are inherently destabilizing since if unchecked they would cause voice F_0 to drift further and further from the intended F_0. It is presently unknown what causes "following" responses. Both response types have latencies of approximately 100–150 ms, and magnitudes are generally a fraction of the stimulus. For example, a 100 cents stimulus (100 cents = 1 semitone) generally yields about a 30 cents response. Table I lists the stimulus magnitudes involved in pitch-shift reflex studies in recent years. Although stimulus magnitudes ranged from 25 to 600 cents, 100 cents being the most widely used (Bauer and Larson, 2003; Burnett and Larson, 2002; Donath *et al.*, 2002; Kiran and Larson, 2001; Natke *et al.*, 2003; Natke and Kalveram, 2001; Sivasankar *et al.*, 2005), response magnitudes rarely exceeded 60 cents. Larson *et al.* (2001) reported a response magnitude of 26 to a 25 cent stimulus, but it is the only report of a full compensation to the pitch perturbations. Similarly, Bauer *et al.* (2006) reported a response of 0.99 to a 1 dB loudness perturbation during production of vowels with a low voice amplitude. These two studies suggest that a full response only occurs with small magnitude stimuli. Therefore, one hypothesis tested in the present study is that the reflexive mechanism may respond to perturbations with magnitudes equal to the stimuli only for small variations in voice feedback.

Recent studies of the pitch-shift reflex have examined the task-dependent role of auditory feedback in control of voice F_0 during speech production (Bauer, 2004; Chen *et al.*, 2007; Natke *et al.*, 2003; Xu *et al.*, 2004). For example, significantly larger responses were found in singing compared to a speaking condition (Natke *et al.*, 2003), and in speech compared to a vowel condition (Chen *et al.*, 2007). To date, however, there is no evidence whether there is a task-dependent modulation of pitch-shift response magnitude in sustained vowels. A recent loudness-shift study has demonstrated larger responses when subjects sustained vowels using a soft voice compared to a normal voice amplitude (Bauer *et al.*, 2006). In an analogous way, voice F_0 level may have a similar effect on responses to pitch-shifted voice feedback. Compared to a low F_0 level, vocalizing at a high F_0 requires greater degrees of muscle contraction (Hirano *et al.*, 1970; Titze, 1994), which may affect response magnitudes or

[a)]Author to whom correspondence should be addressed. Electronic mail: clarson@northwestern.edu

TABLE I. Summary of the studies that compare the stimulus magnitude in vowel phonations.

Study	Stimulus magnitude (cents)	Average response magnitude (cents)
Burnett *et al.*, 1998	25, 50,100,150,200,250,300	3–100
Larson *et al.*, 2001	25,100, 200	26–48
Chen *et al.*, 2007	50,100, 200	22–32
Larson *et al.*, 2007	50	10–18
Natke *et al.*, 2003	100	47–66
Kiran and Larson, 2001	100	30–72
Burnett and Larson, 2002	100	18–24
Bauer and Larson, 2003	100	12–15
Hain *et al.*, 2000	100	20–40
Donath *et al.*, 2002	100	50
Sivasankar *et al.*, 2005	100	9–23
Natke and Kalveram, 2001	100, 600	15–65

latencies to perturbations in voice pitch feedback. That is, a reflexive input to motor neurons that are discharging at a high rate may lead to a greater level of muscle contraction (and hence greater change in F_0) than an equal input to neurons discharging at a lower rate. Therefore, another hypothesis tested was whether response magnitudes to pitch-shifted voice feedback would be larger when subjects vocalized at a high F_0 level compared to a low F_0 level.

Additionally, in previous studies subjects were instructed to phonate at a "comfortable" F_0 or loudness level during sustained vowels or speech production. Brown *et al.* (1976) noted, however, that comfortable level for both frequency and intensity can vary markedly across the experimental sessions. Through a trial period of five successive days, researchers found that subjects varied voice F_0 as much as 30 Hz and amplitude by 25.3 dB during sustained vowels and phrases at a comfortable effort level. Such variation in vocalization could lead to greater variability in response measures. Therefore, in the present study, the subjects were instructed to vocalize the vowel /u/ to match a piano note with a constant F_0 level (high or low) at approximately 70 dB. It was anticipated that this control would lead to reduced variability in response measures and hence increased accuracy for the assessment of the effects of stimulus magnitude and F_0 level on responses to pitch-shift voice feedback.

The third independent variable manipulated in this study was stimulus direction: upwards or downwards shifts in voice pitch feedback. Stimulus direction was not hypothesized to have an effect on the responses during vowel phonations. However, in this experimental paradigm, by randomly altering stimulus direction, it further reduces the chances that the subjects would be accustomed to the stimulus direction. Such an expectation could conceivably affect the responses.

II. METHODS

A. Subjects

Twenty-two subjects (five males and 17 females; ages 19–28), most of whom were students at Northwestern University, participated in the experiment. All subjects passed a hearing screening at 25 dB hearing level bilaterally at 250, 500, 1000, 2000, and 4000 Hz, and none reported a history of neurological or communication disorders. All signed informed consent approved by the Northwestern University Institutional Review Board and were paid for their participation.

B. Apparatus

Subjects wore Sennheiser headphones with attached microphone (model HMD 280) in a sound-treated room throughout the testing. They were asked to vocalize the vowel /u/ at approximately 70 dB sound pressure level (SPL), self-monitoring their voice loudness from a Dorrough Loudness Monitor (model 40-A) placed 0.5 m in front of them. The vocal signal from the microphone was amplified with a Mackie mixer (model 1202) and shifted in pitch with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise (Goldline Audio Noise Source, model PN2; spectral frequencies 1–5000 Hz) with a Mackie mixer (model 1202-VZL), and then amplified with a Crown D75 amplifier and HP 350 dB attenuators at 80 dB SPL. MIDI software (Max/MSP v.4.1 by Cycling 74) was used to control the harmonizer. A Brüel and Kjær sound level meter (model 2250) and in-ear microphones (model 4100) were used to calibrate the microphone and headphones to make sure there was a gain of 10 dB SPL between the subject's voice amplitude and the feedback loudness. The voice output, feedback and TTL control pulses were digitized at 10 kHz, low-pass filtered at 5 kHz, and recorded using Chart software (AD Instruments). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc.). A keyboard (Yamaha, PSR-310) was used to present musical notes through the headphones to the subject prior to each set of trials.

C. Procedures

Before the experiment, subjects were first instructed and tested on their ability to produce a /u/ vowel and match the pitch of two different piano notes from the keyboard. One note was close to average values of conversational F_0 and the other was a much higher level. These notes were based on the gender of the subject as well as on testing on the subjects to make sure they were in a comfortable range for the subjects: for male subjects, C4 (261.83 Hz) was used for the high F_0 and C3 (130.81 Hz) for the low F_0; for female subjects, E4 (329.63 Hz) was used for the high F_0 and A3 (220 Hz) for the low F_0 level. Several practice trials were given to make sure the subjects were able to match the notes within 100 cents. Then the voice was recorded and the F_0 was measured with a fast Fourier transform algorithm (Chart software) to verify that the subjects matched the note. If their voice F_0 was higher than the piano note, they were asked to reduce the voice F_0 and tested again, and vice versa. Two subjects were excluded from the data analysis because one was unable to match the note with high F_0, and the other was unable to keep the voice F_0 constant. Therefore, data were analyzed from 20 subjects.

After the training, the subjects were instructed to repeatedly sustain the vowel /u/ for approximately 5 s duration at either the high or low F_0 level. Before each vocalization, the MIDI program automatically presented the piano note (0.5 s duration) through the headphones. The intensity of the piano notes was not calibrated, but was judged to be at a comfortable loudness level by the experimenters. Subjects were requested to vocalize the vowel /u/ to match the note during the experiment. Production of 12 consecutive vocalizations constituted an experimental block, and all trials within one block consisted of the same piano note. For each vocalization within a block, the voice pitch feedback was increased, decreased or held constant (no stimulus) five times in succession (randomized sequence), resulting in 20 increasing, 20 decreasing, and 20 control perturbations in each block of 12 trials. During each vocalization, the inter-stimulus interval varied between 900 and 1500 ms. The duration of each stimulus was 200 ms and the magnitude was held constant at ±10, 20, 30, 40, or 50 cents within each block. Since the initial pitch perturbation occurred between 500 and 850 ms after vocal onset and the sequence and timing of subsequent stimuli were randomized, subjects could not predict the timing or direction of stimuli that would occur on any given trial.

D. Data analysis

The voice wave form was processed in Praat (www.Praat.org) using an autocorrelation method to produce pulses for each glottal cycle in the vocalizations. This process was done separately for each subject and each experimental condition. These signals were then transformed in Igor Pro to produce a F_0 contour for each vocalization. Event-related averaging techniques were used to measure the voice F_0 response separately to each stimulus direction or for the control trials. The averaging process was done by triggering the averaging program on one of three TTL pulses generated by the MIDI program at the time of the recording. The averaging window included a 200 ms pre and 700 ms poststimulus period. The program then calculated an average wave form for each stimulus direction, or control condition, and for each experimental condition and subject. By subtracting the average wave of the control trials from the average wave for either increasing or decreasing stimulus test waves, a "difference wave" was calculated for upward or downward stimulus trials. A point-by-point series of t tests was run between all control and all test trials, in which response latency was defined as the point where the "p" values of significant differences dipped below 0.02 following the onset of the stimulus with a delay of at least 60 ms and remained significant for at least 50 ms. Those cases where the p wave failed to reach a value of at least 0.02 for at least 50 ms, commencing at least 60 ms after the stimulus, were labeled as nonresponses. Response magnitude was measured as the greatest value of the difference wave following the latency and before the time where the p wave re-crossed the 0.02 value indicating the end of the response. Also, percent response magnitude was calculated by dividing the response magnitude by the stimulus magnitude. Voice jitter measures were made in Praat (local jitter) on all the data files for each subject, which is expressed as the average absolute difference between consecutive periods, divided by the average period. Response magnitude, latency, percent response magnitude and voice jitter were submitted to significance testing using repeated-measures analyses of variance (ANOVAs) (SPSS, v. 11.0). Responses to upward and downward stimuli were not tested separately but were averaged together in the statistical analyses. Nonresponses were replaced by the mean value calculated from the measured data from other subjects for that condition. A log transformation was done on response magnitude, latency and percent magnitude measures to achieve a normal distribution and homogeneity of variance. Assumptions of compound symmetry and circularity for a repeated measures ANOVA were met. Counts of opposing and "following" responses were made under each condition, and a Chi-square analysis was done to assess significance of different counts across conditions. An alpha level of $p < 0.05$ was considered to be statistically significant.

TABLE II. Total number of following (FOL), opposing (OPP), and nonresponse (NR) across F_0 level.

	High	Low	Total
FOL	7	17	24
OPP	191	177	368
NR	2	6	8
Total	200	200	400

III. RESULTS

From 20 subjects across five stimulus magnitudes and two F_0 levels, there were 400 possible responses ($20 \times 5 \times 2 \times 2$). Ninety-two percent of the responses "opposed" the stimulus direction. Only 2% of the responses did not meet our criteria of validity and were declared to be nonresponses. Chi-Square tests revealed a greater number of "following" responses in the low F_0 condition than the high F_0 condition ($\chi^2 = 4.167, df = 1, p = 0.041$) (see Table II), and no statistically significant differences in response types were found across stimulus direction and stimulus magnitudes.

Figures 1 and 2 show examples of the average responses to pitch-shifted feedback across five stimulus magnitudes and two stimulus directions for high and low F_0 levels. All traces have been de-meaned in order to illustrate magnitudes

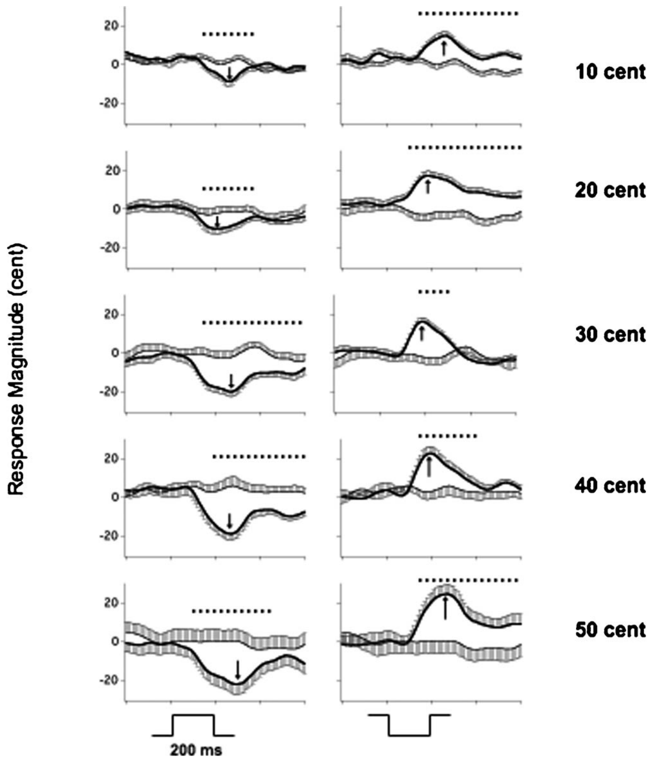

FIG. 1. Control (thin black line) and test average waves (thick black line) for stimulus magnitudes of 10, 20, 30, 40, and 50 cents at a high F_0 level for one subject. The vertical arrow indicates time where the response magnitude was measured. Error bars represent the standard error of the mean for a single direction. Dotted lines at top indicate time of significant difference between test and control waves. Vertical dashed lines indicate response onset. Square brackets at the bottom indicate the time and the direction of the stimulus. Voice F_0 values have been de-meaned.

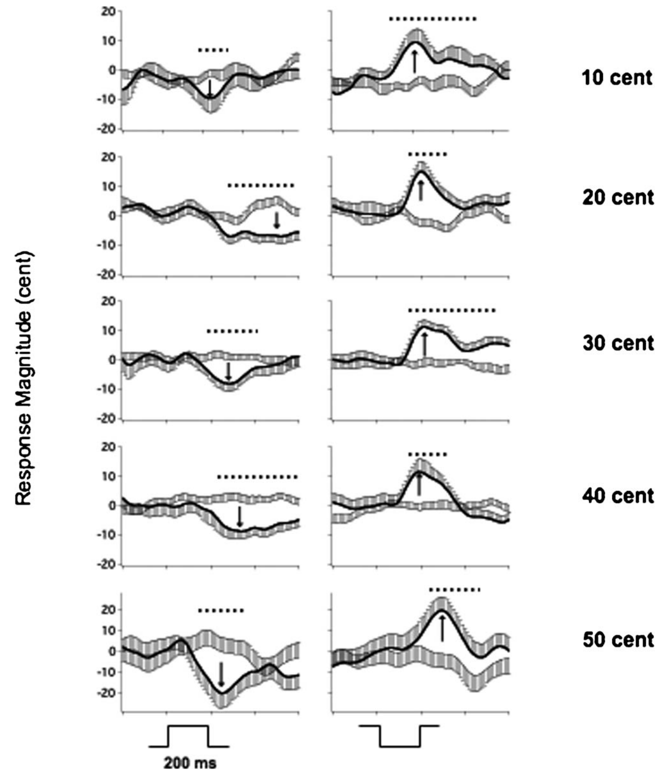

FIG. 2. Control (thin black line) and test average waves (thick black line) for stimulus magnitudes of 10, 20, 30, 40, and 50 cents at a low F_0 level for the same subject as in Fig. 1.

on the same scale. Responses to the upward stimuli are illustrated on the left and downward stimuli on the right, and the responses to ±10, 20, 30, 40, and 50 cents are displayed from the top to the bottom. For each graph, the heavy line represents the average of the responses to the pitch perturbation and the light line the control responses. In all the plots of these two figures, the responses to the pitch perturbation are in the opposite direction to the stimulus. It is also clear, whether for the high or low F_0 condition, that F_0 contours of the control waves are relatively constant. It is difficult to see patterns in the responses in these figures; however, it appears that responses to downward stimuli may be somewhat larger than those to upward stimuli. It also appears that most responses began less than 200 ms after stimulus onset.

Figures 3–5 show boxplots of response magnitudes, response latencies, and percent response magnitude, respectively, as a function of stimulus magnitude for high and low F_0 levels. Values of response magnitude and latency are shown in Tables III and IV. Figure 3 illustrates the increase in response magnitude with stimulus magnitude, while an opposite effect is seen when the same data are plotted as percent response magnitude shown in Fig. 5. The percent response magnitudes were greatest with 10 cents stimuli ($\approx 100\%$) and decreased with greater stimulus magnitudes (Table III). A two-way repeated-measures ANOVA (stimulus magnitude and F_0 level) performed on the log transformed measures of response magnitude indicated significant main effects for stimulus magnitude ($F(4,76)=43.989, p<0.001$) and F_0 level ($F(1,19)=7.318, p<0.002$). There was no significant interaction between stimulus magnitude and F_0 level ($F(4,76)=1.864, p>0.1$). Posthoc Bonferroni tests indicated that 40(16.93 ± 7.72 cents) and 50 cents(18.48 ± 8.29 cents) conditions led to larger responses than 10(9.06 ± 3.88 cents), 20(12.27 ± 4.82 cents), and 30 cents stimuli (14.55 ± 6.65 cents)($p<0.04$) conditions (Fig. 3). No significant differ-

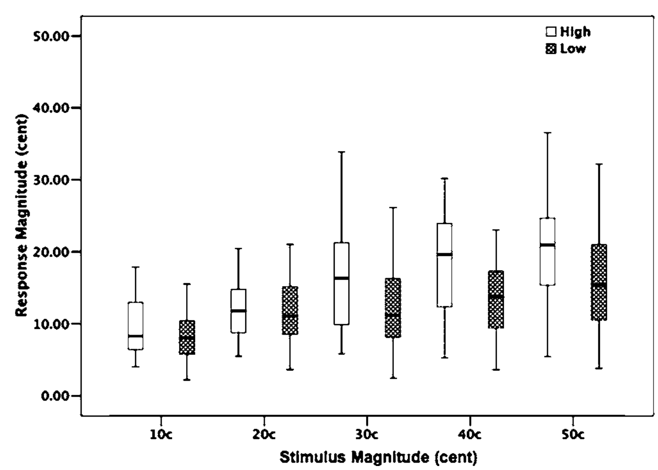

FIG. 3. Box plots illustrating the absolute response magnitude as a function of stimulus magnitude for the high F_0 and low F_0 levels. Shaded boxes are responses for low F_0 level and open boxes for high F_0 level. Box plot definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge – low hinge), and low hinge −1.5 (high hinge–low hinge).

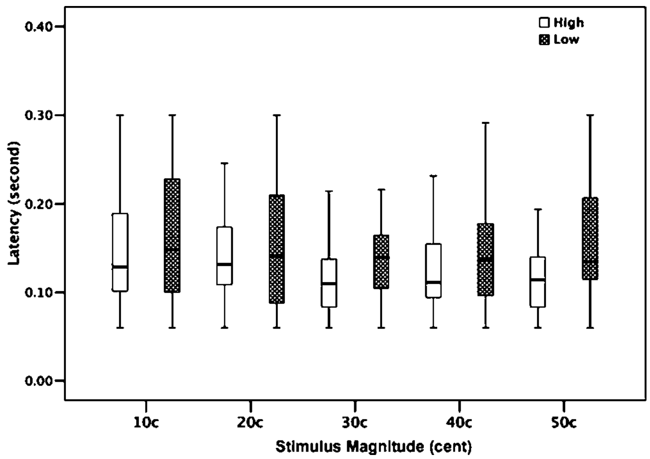

FIG. 4. Box plots illustrating the response latencies as a function of stimulus magnitude for the high F_0 and low F_0 levels.

ences were found between 20 and 30 cents stimuli ($p > 0.2$) and between 40 and 50 cents($p > 0.9$).

For the latency, statistical analyses also revealed significant main effects for stimulus magnitude ($F(4,76) = 3.936, p < 0.01$) and F_0 level ($F(1,19) = 26.668, p < 0.001$). There was no significant interaction in the latency between stimulus magnitude and F_0 level ($F(4,76) = 1.247, p > 0.25$). Posthoc Bonferroni tests indicated that the 10 cents(157±76 ms) condition produced a significantly longer latency than the 30 cents(128±52 ms) condition ($p < 0.002$) (Fig. 4).

Statistical analysis of percent response magnitude also revealed significant main effects for stimulus magnitude ($F(4,76) = 78.508, p < 0.001$) and F_0 level ($F(1,19) = 7.471, p < 0.02$). Posthoc Bonferroni tests indicated significant differences between each stimulus magnitude except between 30 and 40 cents (Table V and Fig. 5). No significant interaction was found in the percent response magnitude between stimulus magnitude and F_0 level ($F(4,76) = 1.846, p > 0.1$).

A two-way repeated-measures ANOVA of voice jitter revealed significant main effects for F_0 level ($F(1,19) = 13.993, p < 0.002$) but not for stimulus magnitude

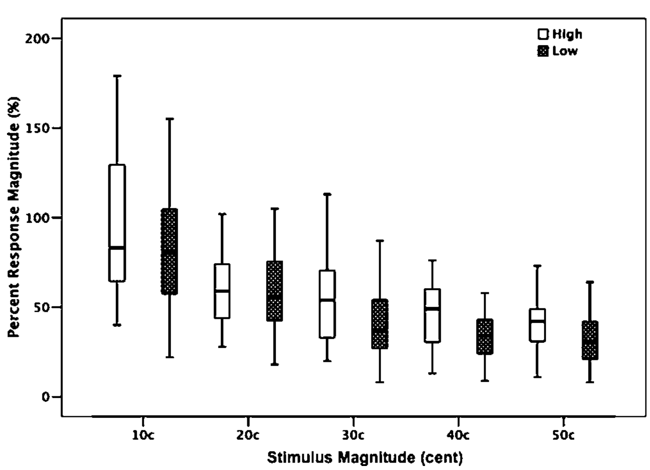

FIG. 5. Box plots illustrating the absolute percent response magnitude as a function of stimulus magnitude for the high F_0 and low F_0 levels.

TABLE III. Average response magnitudes in cents (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	9.66 (4.18)	8.36 (3.44)
20c	12.46 (5.01)	12.08 (4.67)
30c	16.38 (6.94)	12.60 (5.79)
40c	19.24 (7.78)	14.49 (6.97)
50c	21.02 (8.38)	15.80 (7.40)

($F(4,76) = 1.232, p > 0.3$). Higher jitter values were observed for the low F_0(0.33%±0.12%) compared with the high F_0 level (0.25%±0.09%). No significant interaction was found between stimulus magnitude and F_0 level ($F(4,76) = 0.151, p > 0.9$).

IV. DISCUSSION

The purpose of the present study was to investigate vocal responses to small pitch-shift stimuli as a function of vocal F_0 level and pitch-shift magnitude. It was found that response magnitudes were about equal to the 10 cents stimuli, but when expressed as a percent of stimulus magnitude, they decreased with larger stimulus magnitudes. We also found that larger responses were associated a high vocal F_0 level than with the low F_0 condition. In addition, response latencies were shorter for 30 cents compared with 10 cents stimulus magnitudes and with the higher vocal F_0 level compared to the low F_0 level. Thus, greater stimulus magnitudes or a high vocal F_0 level led generally to larger responses and shorter latencies.

The results also showed that, in the case of pitch perturbations as small as 10 cents, the audio-vocal system is capable of compensating for errors in the voice F_0 output. Due to the nonlinear relationship between cent and frequency, the pitch perturbations expressed in Hertz vary across the vocal F_0 level as follows:

$$\Delta F = \left(10^{\frac{\Delta \text{cent}}{3986}} - 1\right) \times F_0,$$

where ΔF is pitch shift in Hertz and F_0 is voice frequency in Hertz. Corresponding to the pitch changes from 10 to 50 cents, the frequency perturbations ranged from 0.76 to 3.85 Hz for males and from 1.28 to 6.47 Hz for females in the low F_0 condition, and from 1.52 to 7.70 Hz for males and 1.92 to 9.70 Hz for females in the high F_0 condition. Hence, one question that arises is whether the small pitch perturbations can be perceived by the auditory system. The minimum detectable change in frequency, or pitch discrimination level

TABLE IV. Average response latencies in ms (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	145(68)	171(83)
20c	140(56)	145(67)
30c	115(41)	142(59)
40c	133(63)	154(80)
50c	121(52)	154(67)

TABLE V. Average percent response magnitude (SD) across stimulus magnitude and F_0 level.

	High F_0	Low F_0
10c	97(41)	84(34)
20c	62(25)	60(23)
30c	55(23)	42(19)
40c	48(19)	36(17)
50c	42(17)	32(15)

for pure tones, was reported as 1.0 Hz(8.6 cents) for both 200 Hz and for 400 Hz(4.3 cents) at a sensation level (SL) of 40 dB, and 1.2 Hz(5.2 cents) for 400 Hz at 80 dB SL-(Wier *et al.*, 1977). Harris (1952) reported that at 40 dB SL the pitch discrimination threshold was 0.4 Hz(5.5 cents) for 125 and 0.75 Hz(5.2 cents) for 250 Hz, respectively. In general, pitch discrimination levels decreased with increases in sensation level regardless of frequency (Harris, 1952; Wier *et al.*, 1977). It is somewhat problematic to make a direct comparison between results of psychophysical testing on pure tones and the complex vocal signals in the present study, but assuming some correspondence between the two techniques, it appears that the 10 cents stimulus magnitudes in this study were somewhat greater than the minimal detectable change in frequency for a pure tone stimulus as reported by others. However, given that the 0.2 s pitch-shift stimuli were shorter in duration than those used by Wier *et al.* (1977) (0.5 s duration) or Harris (1952) (1.4 s duration), the pitch-shift stimuli used in this study may have been rather close to the threshold of detection.

In previous studies, it was shown that response magnitudes rarely exceeded 60 cents in the presence of 50, 100 or 200 cents stimuli (Bauer, 2004; Bauer and Larson, 2003; Burnett *et al.*, 1998; Chen *et al.*, 2007; Hain *et al.*, 2000; Kawahara, 1995; Kiran and Larson, 2001; Larson *et al.*, 2001; Sivasankar *et al.*, 2005). In such cases, percent response magnitudes were less than 50%, varying from approximately 10% to 40%. Larson *et al.* (2001) reported a mean response magnitude of 26 cents in response to 25 cents stimuli, where percent response magnitude was over 100%. In the present study, the mean response magnitudes were also about equal to the 10 cents stimuli (see Table III) but failed to equal the pitch perturbation for stimuli larger than 20 cents. Similar findings were also found in the loudness-shift study of sustained vowels (Bauer *et al.*, 2006), in which the response magnitude to 1 dB loudness perturbation was 0.99 dB at a soft amplitude level, but full compensation was not achieved for 3 or 6 dB amplitude stimuli.

Collectively, these observations indicate that the audio-vocal system can regulate voice F_0 or amplitude with a response magnitude equal to small perturbations, i.e., 10 cents or 1 dB. The failure of the system to correct for errors of larger magnitudes may be a self-protection mechanism to prevent environmental sounds from exerting a predominant influence over the voice output. If the audio-vocal system had a very high gain, such as the oculomotor in which a high gain is required to keep images stabilized on the retina (Glimcher, 1999), voice feedback or environmental sounds could exert greater control over voice F_0 than voluntary mechanisms. Such mechanisms probably explain why even small stimuli may not always produce full compensation. For example, response magnitude of less than 14 cents was produced in response to 25 cents stimuli (Burnett *et al.*, 1998). Also, as Bauer *et al.* (2006) reported for loudness-shifted voice feedback, a gain of close to 1 was only reported when subjects attempted to maintain a relatively quiet voice amplitude, not when they vocalized at a louder amplitude. Hence, a full response to small stimuli may be related to mechanisms involved in the control of specific vocal conditions. To date, there is no evidence of a full response to pitch perturbations during speech production.

The finding that larger response magnitudes occurred with a high F_0 level compared to the low F_0 condition supports the hypothesis that task-dependent modulation of these responses occurs with vowels. This observation is analogous to the study in which it was shown that voice amplitude affected responses to loudness-shifted feedback (Bauer *et al.*, 2006). One explanation for this observation is that maintaining a relatively high F_0 that is clearly beyond the typical conversational level may require greater reliance on auditory feedback than a lower F_0 level. Another explanation is that response magnitude may be related to the phonation stability of sustained vowels (cycle-to-cycle variations in frequency). Previous studies of F_0 effects on the phonation stability have shown that pitch perturbation values (jitter) are greater in low F_0 conditions than in a high F_0 condition (Gelfer, 1995; Horii, 1979). In the present study it was also found that voice jitter was greater in the low F_0 condition compared to high F_0, perhaps making the pitch-shift stimuli more salient in the high F_0 condition than in the low F_0 condition, which may have led to larger and more clearly defined responses.

It is also possible that the greater voice variability (jitter) in the low F_0 condition may be related to the greater number of "following" responses in this condition. Although the cause of "following" responses has not been identified, it was suggested (Hain *et al.*, 2000) that they are a result of the subject treating the feedback signal itself as the referent, as when a singer attempts to match a piano note. In such a case, the singer would adjust their F_0 towards the piano note (external referent). Alternatively, when a singer attempts to produce a remembered pitch, and if their voice pitch feedback does not agree with the memory (internal referent), they adjust their F_0 away from the feedback and towards the memory. In a more recent study (Larson *et al.*, 2007), greater numbers of "following" F_0 responses were observed with stimuli consisting of simultaneous changes in pitch and loudness feedback when the stimuli changed in opposite directions. It was suggested that the following responses were due to difficulty in correctly identifying the stimulus direction. Therefore, in the present study, the greater numbers of "following" responses in the low F_0 condition may have been caused by the greater difficulty in identifying the pitch-shift direction in the midst of a variable voice. It is noteworthy that some people are better than others at detecting the direction of pitch changes (Semal and Demany, 2006), which may explain why the numbers of "following" responses in this and previous studies do not follow a predictable pattern.

It is also possible that vocalizing in the high F_0 condition was a more difficult vocal task, which required greater attention to auditory feedback for regulation. Such task dependency has been reported in previous studies of the effects of perturbed feedback on voice control (Bauer *et al.*, 2006; Chen *et al.*, 2007; Natke *et al.*, 2003; Xu *et al.*, 2004).

Previous studies have demonstrated that response latency can be modulated as a function of stimulus magnitude (Larson *et al.*, 2001) and vocal task (Burnett and Larson, 2002; Chen *et al.*, 2007). The current findings showed longer latencies for the 10 cents stimuli, but with larger stimulus magnitudes, the latencies were shorter. Shorter latencies were also observed with the higher F_0 level. The shorter latencies observed with the greater stimulus magnitudes may reflect the fact that these stimuli were more easily perceived. Similar findings have been reported in the reaction time literature (Jaskowski *et al.*, 1994; Seitz and Rakerd, 1997). The finding that latencies decreased as vocal F_0 level increased is similar to the findings from pitch perception studies where it has been found that subjects are more sensitive to pitch deviations at a high frequency compared to a low frequency (Harris, 1952; Wier *et al.*, 1977). Thus the latency changes observed in the present study may reflect more general characteristics of the interaction between stimulus salience and the speed of response.

V. CONCLUSION

The results of the present study show that responses to perturbations in voice pitch feedback increase in magnitude as the pitch-shift stimuli increase from 10 to 50 cents. However, when response magnitude is considered as a percent of stimulus magnitude, percent response magnitudes decreased from over 90% for 10 cents stimuli to about 37% for 50 cents stimulus magnitudes. These findings suggest the audio-vocal system is optimally suited for stabilization of the voice around small perturbations. Larger response magnitudes were also recorded when subjects maintained a relatively high F_0 compared to a lower level, which may relate to either the lower voice F_0 variability found at high F_0 compared to low levels or a greater reliance on auditory feedback in the high F_0 condition. Altogether, the significant findings in this study may be due in part to the increased precision in the experimental procedures that required subjects to control their voice F_0 levels by matching a tone that was presented just before each vocalization.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. 1R01DC006243. The authors thank Chun Liang Chan for programming assistance and Kaavya Paruchuri for assistance in data recording. They authors also thank two anonymous reviewers who made important suggestions that improved ithe manuscript.

Bauer, J. J. (**2004**). "Task dependent modulation of voice *F*0 responses elicited by perturbations in pitch of auditory feedback during English speech and sustained vowels," Ph.D. Dissertation, Northwestern University, Evanston, Il.

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.

Brown, W., Murry, T., and Hughes, D. (**1976**). "Comfortable effort level: An experimental variable," J. Acoust. Soc. Am. **60**, 696–699.

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.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F_0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**, 1157–1163.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Gelfer, M. P. (**1995**). "Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability," J. Speech Hear. Res. **38**, 1189–1198.

Glimcher, P. W. (**1999**). "Eye movements," in *Fundamental Neuroscience*, edited by M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts and L. R. Squire (Academic, San Diego), pp. 993–1010.

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.

Harris, J. D. (**1952**). "Pitch discrimination," J. Acoust. Soc. Am. **24**, 750–755.

Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch and intensity of voice," Folia Phoniatr. **22**, 1–20.

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

Jaskowski, P., Rybarczyk, K., and Jaroszyk, F. (**1994**). "The relationship between latency of auditory evoked potentials, simple reaction time, and stimulus intensity," Psychol. Res. **56**, 59–65.

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.

Kiran, S., and Larson, C. R. (**2001**). "Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's Disease patients and normal controls," J. Speech Lang. Hear. Res. **44**, 975–987.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparisons of voice F_0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.

Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control," J. Acoust. Soc. Am. **121**, 2862–2872.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.

Seitz, P. F., and Rakerd, B. (**1997**). "Auditory stimulus intensity and reaction time in listeners with longstanding sensorineural hearing loss," Ear Hear. **18**, 502–512.

Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.

Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.

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.