Voice F_0 responses to pitch-shifted voice feedback during English speech

Stephanie H. Chen
Feinberg School of Medicine, Northwestern University, 440 North McClurg Ct. #604, Chicago, Illinois 60611

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

Yi Xu
Department of Phonetics and Linguistics, University College London, London, United Kingdom

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

(Received 1 June 2006; revised 3 November 2006; accepted 8 November 2006)

Previous studies have demonstrated that motor control of segmental features of speech rely to some extent on sensory feedback. Control of voice fundamental frequency (F_0) has been shown to be modulated by perturbations in voice pitch feedback during various phonatory tasks and in Mandarin speech. The present study was designed to determine if voice F_0 is modulated in a task-dependent manner during production of suprasegmental features of English speech. English speakers received pitch-modulated voice feedback (±50, 100, and 200 cents, 200 ms duration) during a sustained vowel task and a speech task. Response magnitudes during speech (mean 31.5 cents) were larger than during the vowels (mean 21.6 cents), response magnitudes increased as a function of stimulus magnitude during speech but not vowels, and responses to downward pitch-shift stimuli were larger than those to upward stimuli. Response latencies were shorter in speech (mean 122 ms) compared to vowels (mean 154 ms). These findings support previous research suggesting the audio vocal system is involved in the control of suprasegmental features of English speech by correcting for errors between voice pitch feedback and the desired F_0. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2404624]

PACS number(s): 43.70.Mn, 43.72.Dv, 43.70.Bk [AL] Pages: 1157–1163

I. INTRODUCTION

Little is known about neural mechanisms controlling voice fundamental frequency (F_0) during speech. In English and other nontonal languages, F_0, along with amplitude and duration, are all increased for stressed syllables and at the end of a phrase or sentence to indicate a question (Alain, 1993; Cooper *et al.*, 1985; Eady and Cooper, 1986; Lieberman, 1960; Xu and Xu, 2005). F_0 is thus important in the overall goal of speech communication and to convey emotional expression (Bänziger and Scherer, 2005; Chuenwattanapranithi *et al.*, 2006). In some types of neurologically based voice disorders, voice F_0 is often abnormal and interferes with communication (Duffy, 1995). Understanding mechanisms of F_0 control during speech is important for treatment and prevention of some types of voice disorders.

Theoretical discussions of speech motor control in the past have focused primarily on segmental features of speech. To this end, suggestions have been advanced that segmental features may be controlled by an internal model or a motor plan guided in part by sensory feedback (Fairbanks, 1954; Gracco and Abbs, 1985; Munhall *et al.*, 1994). Several studies in recent years have also demonstrated through the use of the perturbation paradigm that auditory feedback is important for the on-line control of voice F_0 during sustained vowels (Bauer and Larson, 2003; Hain *et al.*, 2000; Larson *et al.*, 2001; Sivasankar *et al.*, 2005), glissandos (Burnett and Larson, 2002), singing (Natke *et al.*, 2003), nonsense syllables produced by German speakers (Donath *et al.*, 2002; Natke *et al.*, 2003; Natke and Kalveram, 2001), and during prolonged vowels in the context of Mandarin phrases (Jones and Munhall, 2002). A simple mathematical model based on negative feedback accounts for the main features of these responses (Bauer *et al.*, 2006; Hain *et al.*, 2000). It was also found in normal Mandarin speech that the magnitudes of F_0 responses to pitch perturbations were larger in phrases in which there was a subsequent fall in F_0 (high-falling or high-rising phrases) compared to a phrase where the F_0 remained relatively constant (high-high phrase). These observations suggest that there is task-dependent modulation of pitch-shift responses in Mandarin (Xu *et al.*, 2004a). Similarly, Natke *et al.* (2003) provided evidence that pitch-shift responses are modulated according to the demands of the vocal task by showing that responses to pitch perturbations were larger in

[a)]Electronic mail: clarson@northwestern.edu singing compared to speaking nonsense syllables. To date, no studies have demonstrated whether there is task-dependent modulation of the pitch-shift response magnitude in a nontonal language such as English.

In a recent study of normal English speech, it was found that perturbations in voice pitch auditory feedback led to changes in the timing of suprasegmental features (Bauer, 2004). When the direction of the pitch-shift stimulus (e.g., down) was opposite to that of the F_0 change in direction for the inflected syllable (e.g., up), the timing of the peak in the F_0 contour for the inflected syllable was delayed. When the direction of the shift was in the same direction as the inflected syllable, there was no delay. Along with this timing change, response latencies to the pitch-shifted feedback were modulated so as to occur during the peak of the inflection. Possible changes in response magnitude were obscured by the relatively large variations in F_0 corresponding to the suprasegmental features of the sentence.

The present study was designed to explicitly test whether the magnitudes and latencies of responses to pitch-shifted voice feedback are modulated during English speech by using a phrase that did not have the very large variations in the F_0 contour as reported by Bauer (2004). In the present study, subjects were instructed to repeat a phrase in which the F_0 contour was relatively flat and then rose at the very end, as in a question. It was hypothesized that responses to pitch-shifted voice feedback that were presented during speech would be larger than those presented during a sustained vowel task because control of F_0 during speech is important for conveying information to the listener, while control of F_0 during a sustained vowel has no such goal and hence is inherently less meaningful than during speech. Results confirmed that responses to pitch-shifted feedback during speech were larger and faster than those produced during a sustained vowel task.

II. METHODS

A. Subjects

Twenty subjects (10 males and 10 females) between the ages of 19 and 21 were recruited. All subjects reported that English was the first language they learned. All subjects reported normal hearing, and none reported a history of speech or language problems or neurological disorder. All subjects signed informed consent approved by the Northwestern University Institutional Review Board.

B. Apparatus

Subjects were seated in a sound-attenuated chamber for the testing. Sennheiser headphones (model HMD 280) with an attached microphone were placed on the subject. The microphone signal was amplified (Mackie mixer model 1202), shifted in pitch with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL masking noise (low-pass filtered from 10 to 5000 Hz), and then amplified to 10 dB SPL greater at the Sennheiser headphones than at the microphone. Subjects monitored their voice amplitude on a Dorrough Loudness monitor (located 0.5 m in front of the subject) in an attempt to keep their vocal level near 70 dB SPL. Voice, feedback, and TTL control pulses (generated by a locally fabricated circuit and controlled by MIDI software) were digitized at 10 kHz (5000 Hz low pass filter) on a laboratory computer. Acoustic calibrations were made with a Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100).

C. Procedures

Subjects were first instructed that they would hear a phrase ("you know Nina?") spoken over headphones (female voice), and that they should repeat the phrase within 1 s in exactly the same manner as that of the sample. Because the phrase was spoken as a question, it started with a flat F_0 trajectory and then rose on the final syllable "…na" (Eady and Cooper, 1986).

The MIDI program initiated a trial by first presenting the voice recording to the subject. The onset of the subject's voice then caused the MIDI program to activate the harmonizer and deliver the pitch-shift stimulus to the subject with a delay of 200 ms following voice onset. This delay time was chosen, based on measurements of the model phrase, so that the stimulus and response would begin before the rise in F_0 for the final syllable (na). It was necessary for the response to begin before the rise in F_0 so that we could measure it independently of the rise in F_0 (see the following). There was a 1500 ms intertrial interval. Subjects repeated this task 60 times, which took about 5 min. On one-third of the trials, an upward (increasing pitch) pitch-shift stimulus was presented, on one-third a decreasing pitch-shift stimulus was presented, and on one-third no stimulus (control) was presented. Since in the block of 60 trials, the sequence of stimuli was randomized, subjects could not predict which type of stimulus would occur on any given trial. Across 3 blocks of 60 trials, the stimulus magnitude was varied at ±50, 100, and 200 cents (200 ms duration). Stimulus durations of 200 ms were chosen because longer stimuli elicit voluntary responses by the subject (Burnett *et al.*, 1998).

Subjects were also tested with pitch-shifted voice feedback while repeating sustained vowel phonations. Subjects were instructed to say the vowel /u/ for a duration of approximately 5 s at their conversational pitch level and 70 dB SPL amplitude. During each vocalization, a randomized mixture of five control (no pitch-shift stimulus) or pitch-shift stimuli (±50, 100, or 200 cents) were presented at randomized times. Previous research has shown that this method of testing yields results that are identical to the presentation of one stimulus during each vocalization (Bauer and Larson, 2003). Thus, with each sequence of 12 vocalizations, 60 control or pitch-shift stimuli were presented. During any one block of trials, the pitch-shift magnitude was constant.

For data analysis, the voice, voice feedback, and TTL pulses were digitized at 10 kHz using Chart software (ADInstruments). The voice wave form was then processed in Praat using an autocorrelation method to produce a train of pulses corresponding to the fundamental period of the voice waveform. This pulse train was then converted into an analog wave in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR). The F_0 signals were then converted to a cents scale using the following equation: cents$=100$ $(39.86 \log_{10}(f2/f1))$ where $f1$ equals an arbitrary reference note at 195.997 Hz (G4) and $f2$ equals the voice signal in hertz. The F_0 wave form and TTL pulses were displayed on a computer screen, and the beginning and end points of the F_0 wave for each vocalization were marked with cursors. All the vocalizations in a block of 60 trials were then time-normalized. The time-normalization process was done by first calculating the average duration of all 60 vocalizations in a block of trials and then changing the durations of each of the F_0 traces and the accompanying trace representing the stimulus to the average duration of the entire group. By doing this normalization, the stimulus pulses maintained their alignment with the respective F_0 trace, and the pitch contours of the entire group were aligned in such a way that variability in the averaging process was reduced. Then, the F_0 trace for each vocalization was time-aligned with the TTL pulse representing the pitch-shift stimulus for each trial and an average trace was calculated separately for the two different stimulus directions in a block of 60 trials. An average of the control trials was produced in the same way, only in this case the TTL pulses were not accompanied by any change in voice pitch feedback. Thus, an average F_0 trace was constructed separately for downward stimuli, for upward stimuli and control trials for each subject and for each experimental condition.

After averaging, a statistical test was performed to determine if the average of the control wave differed significantly from the average of the test wave for the upward and downward stimulated trials. A point-by-point series of t-tests were run between all control and all test waves for a given condition and subject (see Xu *et al.*, 2004a). This process yielded an array of "p" values indicating the level of significant difference between the control and test waves. Response latencies were defined as the time point where the p values decreased below 0.02 and remained decreased for at least 50 ms. Rather than using a statistical correction factor to prevent spurious statistical significance from occurring (e.g., Bonferroni correction), we reasoned that physiological criteria provided a more valid approach. It is known from previous studies that a finite time of at least 60 ms occurs between a pitch-shift stimulus and a F_0 response (Burnett *et al.*, 1998; Burnett and Larson, 2002; Burnett *et al.*, 1997; Hain *et al.*, 2000; 2001; Larson 1998; Larson *et al.*, 2001; 1997, 2000). Also, it is known that the fastest contraction speeds of a muscle such as the cricothyroid, which is important for voice F_0 control, is about 30 ms to peak contraction (Perlman and Alipour-Haghighi, 1988), and a change in voice F_0 occurs 20–30 ms later (Kempster *et al.*, 1988; Larson *et al.*, 1987). Therefore, by limiting minimal latencies to 60 ms and response durations to 50 ms, this method guarded against significant changes in the voice F_0 that were not due to activity of the neuromuscular system. If we had employed a correction factor such as Bonferroni, some very short latencies or short duration responses could have been included in the set of acceptable responses.

A "difference wave" was then calculated by subtracting the average control wave from the averaged upward and downward stimulus test waves for each subject and each condition. The difference wave was used to measure the response magnitude, which was measured as the greatest value of the difference wave following the latency and before the time where the p wave recrossed the 0.02 value indicating the end of the response. Response latency and magnitude measures were submitted to significance testing using a repeated-measures ANOVA (SPSS, v. 11.0).

TABLE I. Numbers of following (FOL), nonresponses (NR), and opposing responses (OPP) for the speech and nonspeech vocal conditions.

	Speech	Vowel	Total
FOL	16	5	21
NR	10	2	12
OPP	94	113	207
Total	120	120	240

III. RESULTS

Out of the 240 possible responses, there were 12 nonresponses that did not register a significant difference between the control and test wave, and 10 of these occurred in the speech condition (see Table I). Of the remaining 228, 21 responses were in the "following" direction (the response change in F_0 was in the same direction as the stimulus) and 207 in the opposing direction (response and stimulus waves changed in opposite directions). Sixteen of the "following" responses were in the speech condition, and 5 in the vowel condition. A chi-square test revealed a statistically greater number of nonresponses and "following" responses in the speech condition compared to the vowel condition ($\chi^2 = 12.84$, df$=2$; $p < 0.002$). The distribution of opposing, "following," and nonresponses was even across the upward and downward stimulus directions and the stimulus magnitudes.

Figures 1–3 show examples of responses to pitch-shifted feedback during speech and vowel productions. Figure 1 illustrates responses to a downward 50 cents perturbation in voice pitch feedback on the left and an upward perturbation on the right. The top row shows responses produced during

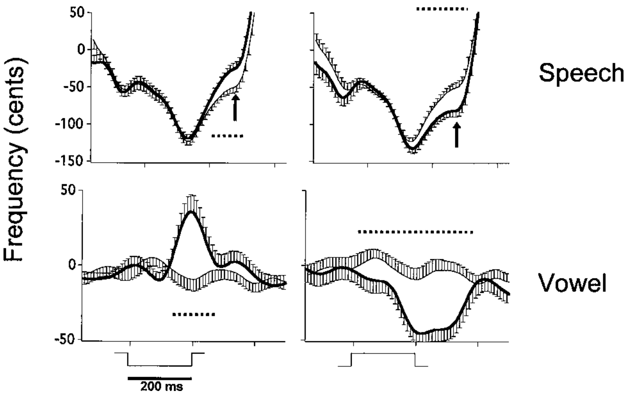

FIG. 1. F_0 contours for speech (top) and vowels (bottom). Traces for downward pitch-shift stimuli are on the left (indicated by square brackets at the bottom) and upward pitch-shift stimuli on the right. Contours with heavy lines are for stimulated trials, and light lines for control trials. Error bars represent 1 s.d. of the mean. Arrows indicate time where response magnitudes for speech contours were measured. Horizontal dashed lines indicate time when the control and test waves differed significantly. Stimulus magnitudes were ±50 cents.

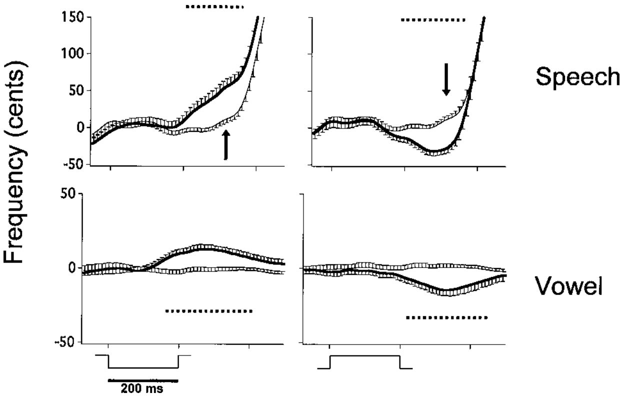

FIG. 2. F_0 contours for speech and vowel productions with ±100 cents stimuli.

speech and the bottom row responses during a vowel. For each graph, the heavy line represents the average of the responses to the pitch perturbation and the light line the control responses. In each case, the responses to the pitch perturbation are in the opposite direction to the stimulus and occur within 200 ms of the onset of the stimulus. It is also clear that the responses during speech began before the rise in the F_0 trajectory of the final syllable and merged with the rising trajectory. In making the measurements of response magnitude, we noted the point of inflection of the F_0 trace that was part of the elevation at the end of the syllable and used this time as a cutoff point for measuring response magnitude from the difference wave. We compared this manual method with local peaks of acceleration of the F_0 trace and found no more than 20 ms disagreement. By making the measurements in this way, we attempted to reduce the likelihood that the response magnitude measures would be exaggerated by the rising F_0 at the end of the phrase. Arrows on the curves in Figs. 1–3 indicate the times where the measures were made. For the examples in Fig. 1, the responses produced during the vowel were larger than those produced during speech. Figures 2 and 3 show similar results for different subjects for the 100 and 200 cents perturbations, respectively. In these examples, as well as those for most of the subjects, all responses produced during speech were larger than those produced during vowels.

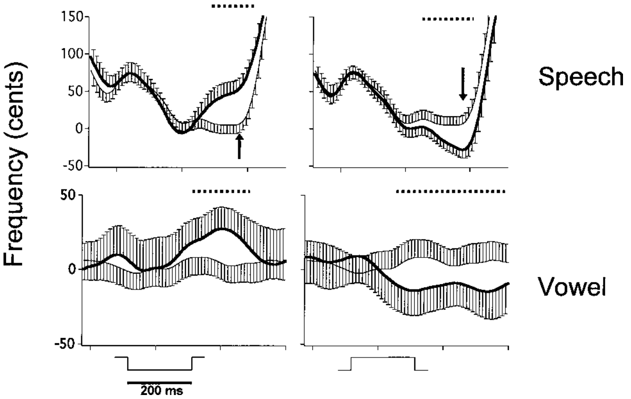

FIG. 3. F_0 contours for speech and vowel productions with ±200 cents stimuli.

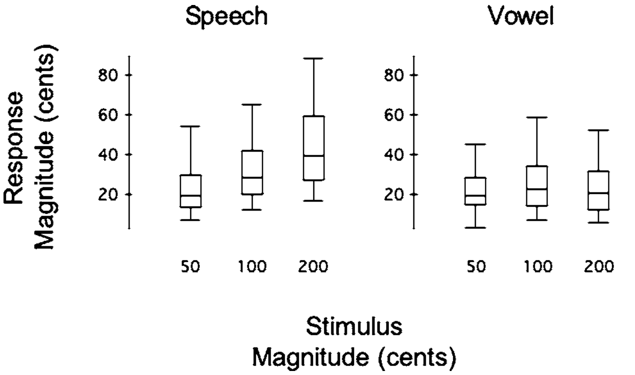

FIG. 4. Boxplots indicating response magnitudes for speech (left) and vowels (right) across stimulus magnitudes of ±50, 100, and 200 cents. Box definitions: Middle line is median, top and bottom of boxes are the 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge-low hinge), and low hinge −1.5 (high hinge-low hinge) (Data Desk; Data Description).

A three-way repeated-measures ANOVA was performed on the response magnitude of opposing responses with task, stimulus magnitude, and stimulus direction. Significant main effects on response magnitude were found for task ($F(1,114)=7.813$, $p<0.01$), stimulus magnitude ($F(2,114)=10.036$, $p<0.001$), and stimulus direction ($F(1,114)=28.332$, $p<0.001$). Responses in the speech condition (mean 31.5 ± 18.7 cents) were larger than those in the vowel condition (mean 21.6 ± 11.7). Significant interactions were between stimulus magnitude and stimulus direction ($F(2,38)=11.330$, $p<0.001$) and between stimulus magnitude and task ($F(2,114)=8.868$, $p<0.001$). Figure 4 shows box plots of response magnitude across the three stimulus magnitudes for all subjects for the speech and vowel conditions. Two-way repeated-measures ANOVAs (stimulus magnitude and stimulus direction) were performed on the magnitude for the speech and vowel conditions, respectively. As can be seen, there was a clear increase in response magnitude as a function of the stimulus magnitude for the speech ($F(2,57)=17.722$, $p<0.001$) but not for the vowel condition. Figure 5 shows box plots of response magnitude for the downward and upward stimuli for both speech and vowels. Here a clear effect of the stimulus magnitude can be seen for the downward stimuli during speech ($F(2,57)=25.996$, $p<0.001$), but not for the upward stimuli. No changes in response magnitudes for the vowels were observed.

A three-way repeated-measures ANOVA was also performed on the response latency with task, stimulus magnitude, and stimulus direction. There was a significant main effect for task, where latencies for speech (mean 122 ± 63 ms) were significantly shorter than those produced during the vowel task (mean 154 ± 79 ms) ($F(1,114)=13.195$, $p<0.0001$). There were no other significant latency effects.

IV. DISCUSSION

The present study was designed to test the hypothesis that the pitch-shift reflex would generate larger responses in an English speech task compared to a nonspeech task. This hypothesis is based on previous observations that reflexive

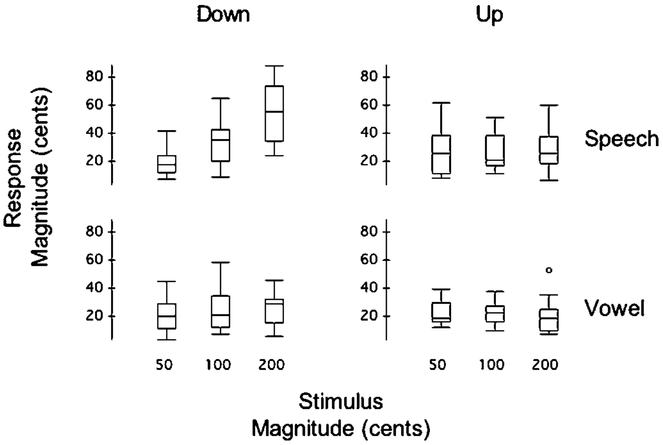

FIG. 5. Boxplots showing response magnitudes for ±50, 100, and 200 cents stimuli separately for downward (left) and upward stimuli (right). Upper row shows results from the speech condition and bottom row for the vowel condition. Box definitions: Middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of the main body of data defined as high hinge +1.5 (high hinge-low hinge), and low hinge −1.5 (high hinge-low hinge) (Data Desk; Data Description).

mechanisms reflect neural connections between sensory feedback and motor control mechanisms (Houk, 1978; Stein, 1980). In other sensorimotor systems, when sensory feedback is important for the successful execution of a task, the controlling mechanisms generate compensatory responses to perturbations in sensory feedback (Gracco and Abbs, 1989; Munhall *et al.*, 1994; Shaiman, 1989; Tremblay *et al.*, 2003). As a subject is performing a motor task, the changes in the execution of the task that are measured in response to sensory perturbation reflect the neural mechanisms that are normally involved in completing the task. In this sense, the value of such studies lies in the knowledge they impart regarding neural mechanisms that control behavior.

In the present study, speakers produced a phrase in which the F_0 was held relatively stable across several syllables, and then elevated at the end of the phrase. This elevation was a suprasegmental adjustment in speech that made the phrase sound like a question. It is one of many ways in which nontonal languages use suprasegmental adjustments to express meaning (Bänziger and Scherer, 2005; Eady and Cooper, 1986; Lieberman, 1960; Xu and Xu, 2005). This particular experimental paradigm was necessitated by the difficulty in eliciting highly consistent intonation patterns from English subjects. In a pilot test, subjects were found to use variable intonation patterns for the sentences. This was probably because the English orthography does not implicitly specify any contrastive pitch patterns, as does the Chinese orthography. This variability could potentially be so large as to make the comparisons impossible. To avoid this problem, we used a prerecorded sentence as a model for the subjects. For each trial, they were asked to speak in the same way as the model sentence. No specific instructions were given, however, as to what aspects of the model they should imitate. This was to guarantee that they would not focus only on maintaining the pitch pattern. A similar imitation paradigm has been used previously in a study of Mandarin intonation (Xu *et al.*, 2004b). Because people's ability to consciously analyze and imitate pitch is highly variable (Dankovicova *et al.*, in press), what is likely involved in performing the task is subjects' linguistic ability rather than their musical ability. Such linguistic ability, though also requiring highly accurate pitch control, involves only the control of relative pitch rather than exact pitch as in singing (Xu, 2005). Thus the larger responses observed in the speech condition compared to the vowel condition may indicate that auditory feedback is used on-line to help control suprasegmental features of speech production.

Additional evidence supporting this conjecture comes from the findings that responses were larger for downward stimuli compared to upward stimuli. The downward pitch perturbation was opposite in direction to the planned upward F_0 trajectory. It is known that questions in English manifest a rising intonation, particularly at the end of the sentence (Bolinger, 1989; Eady and Cooper, 1986; McRoberts *et al.*, 1995; Pell, 2001). Thus the downward pitch-shift would have made it sound to the subject as if his/her voice F_0 was changing in the wrong direction, away from the intended rise. In order to achieve the rising intonation, a greater upward response magnitude would be required compared to nonperturbed (control) trials. Moreover, the fact that the response magnitudes increased along with the magnitude of the downward directed pitch-shift stimuli, indicates that the system not only recognizes errors but is also capable of assessing their relative magnitudes and increasing the magnitude of the compensatory responses. By comparison, upward pitch perturbations, which were in the same direction as the planned inflection pattern, did not interfere with the inflection pattern. In this condition, as well as in the vowel productions, the response magnitudes were smaller than for downward directed stimuli presented during speech, suggesting the need for a corrective response was likely not as great.

Although there were far more opposing than "following" or nonresponses, there were a statistically greater number of "following" and nonresponses in the speech condition compared to the vowel condition. There may be two reasons for this difference. First, during speech, the F_0 trajectories of the control and test trials were dynamically changing, which made it more difficult to measure responses. Because of this, some responses may have been small and did not meet our criteria for acceptance. A second explanation may be that during the speech task, the audio-vocal system misinterpreted the direction of the stimulus and produced a "following" response in error. It is possible that if pitch shift stimuli are small (50–200 cents) and short in duration (200 ms), the system does not always recognize them. Providing larger magnitude and/or longer duration stimuli may reduce the number of nonresponses. It is more difficult to explain the higher percentage of "following" responses since there is still no clear explanation of their cause. However, it is possible that during speech, when F_0 is changing dynamically, subjects may misperceive the direction of the pitch-shift stimuli and respond inappropriately.

In this and previous pitch-shifting (Bauer and Larson, 2003; Burnett *et al.*, 1998; Burnett and Larson, 2002; Donath *et al.*, 2002; Elman, 1981; Hain *et al.*, 2000; Jones and Munhall, 2002; Kawahara, 1995; Natke *et al.*, 2003; Sivasankar *et al.*, 2005; Xu *et al.*, 2004a) and loudness-shifting (Bauer *et al.*, 2006; Heinks-Maldonado and Houde, 2005) studies, and studies of the Lombard response or side-tone amplification (Lane and Tranel, 1971), response magnitudes rarely achieved parity with stimulus magnitude. These findings reveal, as was suggested previously (Burnett *et al.*, 1998), that the audio-vocal system appears to be optimized for fine-tuning of voice F_0 or amplitude. The fact that the system can respond to 25 cent stimuli with a response of 26 cents shows that the system is optimally suited for correcting small errors in voice F_0 output (Larson *et al.*, 2001). It is also known that the system will respond to sounds other than those of the speaker (Sivasankar *et al.*, 2005). If the system responded to acoustical perturbations with responses of the same magnitude as the perturbation itself, environmental sounds could exert a predominant influence over the voice. Instead, by responding only partially to auditory feedback perturbations, the system allows for voluntary and cognitive mechanisms to be the most important factors controlling the voice. Moreover, the fact that response magnitudes can increase in certain speaking or singing conditions (Bauer *et al.*, 2006; Natke *et al.*, 2003; Xu *et al.*, 2004a), reveals the flexible nature of the audio-vocal system.

In comparing this study with previous ones, it was found that both the timing and magnitude of responses to pitch-shifted voice feedback are modulated during speech. In the present study, response latencies were shorter and magnitudes were greater in the speech condition compared to the vowel condition. Xu *et al.* (2004a) found that in Mandarin speech, response latencies were shorter and magnitudes were larger in speech conditions in which the stimulus direction (down) was introduced prior to a planned drop in voice F_0 (high-rising) compared to the condition where the F_0 contour was relatively stable (high-high phrase). Although the present study provides no evidence that response latencies may be differentially modulated according to different speech contexts, the fact that Bauer (2004) reported latencies to be increased in some English speech contexts and not others, suggests that the timing of voice F_0 responses to perturbations in auditory feedback can be modulated according to variations in the suprasegmental patterns of English speech just as with lexical contrasts in Mandarin. It is also important to note that the latency changes in speech are most likely due to the demands of speech since they have not been observed in other studies on sustained vowel productions (Bauer and Larson, 2003; Burnett *et al.*, 1998; Hain *et al.*, 2000). Thus data from this and previous studies indicate that during speech, both the timing and magnitude of responses to auditory feedback can be adjusted. These adjustments depend on the direction of the pitch-shift stimulus and the context of the speech at the time of the perturbation. The lack of such adjustments during vowels may reflect the fact that there is no differential importance to either an increase or decrease in voice pitch feedback.

Thus, experiments in both English and Mandarin have shown that pitch-shift reflex magnitudes and latencies are modulated during speech. Since Mandarin is a tone language and English is a nontonal language, the findings that the modulations in voice F_0 based on perturbations in voice pitch feedback in both languages are similar suggest that the neural mechanisms underlying these responses are similar in tonal and nontonal languages as well as for segmental and suprasegmental features of speech production.

V. CONCLUSION

The present study demonstrated that neural control of voice F_0 during suprasegmental features of speech production is accomplished with the aid of auditory feedback of voice pitch. Moreover, results demonstrate that the control mechanisms are modulated according to task demands. The response magnitudes in speech were larger than in a nonspeech task and for downward pitch-shift stimuli compared to upward stimuli. Since these responses occurred just before the inflection of voice F_0 associated with a question, it is suggested that the mechanisms controlling responses to voice pitch-shifted feedback are sensitive to the planned inflection in voice F_0. As the subject is planning a rise in F_0, a downward perturbation in voice pitch-feedback elicits a response that attempts to prevent F_0 from going in the wrong direction, so that the forthcoming rise in F_0 can be accurately made.

ACKNOWLEDGMENTS

This study was completed as part of an Undergraduate Honors Research Thesis in the Dept. of Communication Sciences and Disorders, Northwestern University. We acknowledge the contribution of Kaavya Paruchuri and Christine Lin for their assistance in data analysis. This study was supported by NIH Grant No. DC006243-01A1.

Alain, C. (**1993**). “The relation among fundamental frequency, intensity, and duration varies with accentuation,” J. Acoust. Soc. Am. **94**, 2434–2436.

Bänziger, T., and Scherer, K. R. (**2005**). “The role of intonation in emotional expressions,” Speech Commun. **46**, 252–267.

Bauer, J. J. (2004). “Task dependent modulation of voice F0 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.

Bolinger, D. (**1989**). *Intonation and Its Uses—Melody in Grammar and Discourse* (Stanford University Press, Stanford, CA).

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.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). “Voice F0 responses to pitch-shifted auditory feedback; A preliminary study,” J. Voice **11**, 202–211.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., and Maneewongvatana, S. (**2006**). “Expressing anger and joy with the size code,” Transactions on Engineering, Computing and Technology **11**, 222–227.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). “Acoustical aspects of contrastive stress in question-answer contexts,” J. Acoust. Soc. Am. **77**, 2142–2156.

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.

Duffy, J. R. (**1995**). *Motor Speech Disorders* (Mosby, St. Louis).

Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**, 402–416.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.

Fairbanks, G. (**1954**). "Systematic research in experimental phonetics. 1. A theory of the speech mechanism as a servosystem," J. Speech Hear. Res. **19**, 133–140.

Gracco, V. L., and Abbs, J. H. (**1985**). "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes," J. Neurophysiol. **54**, 418–432.

Gracco, V. L., and Abbs, J. H. (**1989**). "Sensorimotor characteristics of speech motor sequences," Exp. Brain Res. **75**, 586–598.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). "Effects of delayed auditory feedback (DAF) on the pitch-shift reflex," J. Acoust. Soc. Am. **109**, 2146–2152.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**, 131–137.

Houk, J. C. (**1978**). "Participation of reflex mechanisms and reaction-time processes in the compensatory adjustments to mechanical disturbances," in *Cerebral Motor Control in Man: Long Loop Mechanisms*, edited by J. E. Desmedt (Karger, Basel) Vol. **4**, pp. 193–215.

Dankovicova, A. C., House, J. J., and Jones, K. (In Press). "The Relationship between musical skills, music training, and intonation analysis skills," Lang. & Speech.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

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.

Kempster, G. B., Larson, C. R., and Kistler, M. K. (**1988**). "Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency," J. Voice **2**, 221–229.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of $F0$ control," J. Commun. Disord. **31**, 489–503.

Larson, C. R., 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., Burnett, T. A., Freedland, M. B., and Hain, T. C. (1997). "Voice $F0$ responses to manipulations in pitch feedback stimuli," First International Conference on Voice Physiology and Biomechanics, Evanston, IL.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift onset velocity on voice $F0$ responses," J. Acoust. Soc. Am. **107**, 559–564.

Larson, C. R., Kempster, G. B., and Kistler, M. K. (**1987**). "Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles," J. Speech Hear. Res. **30**, 552–558.

Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **32**, 451–454.

McRoberts, G. W., Studdert-Kennedy, M., and Shankweiler, D. P. (**1995**). "The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation," Percept. Psychophys. **57**, 159–174.

Munhall, K. G., Löqvist, A., and Kelso, J. A. S. (**1994**). "Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip," J. Acoust. Soc. Am. **95**, 3605–3616.

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.

Pell, M. D. (**2001**). "Influence of emotion and focus on prosody in matched statements and questions," J. Acoust. Soc. Am. **109**, 1668–1680.

Perlman, A. L., and Alipour-Haghighi, F. (**1988**). "Comparative study of the physiological properties of the vocalis and cricothyroid muscles," Acta Oto-Laryngol. **105**, 372–378.

Shaiman, S. (**1989**). "Kinematic and electromyographic responses to perturbation of the jaw," J. Acoust. Soc. Am. **86**, 78–88.

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.

Stein, R. B. (**1980**). *Nerve and Muscle* (Plenum, New York).

Tremblay, S., Shiller, D. M., and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.

Xu, Y. (**2005**). "Speech melody as articulatorily implemented communicative functions," Speech Commun. **46**, 220–251.

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004a**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.

Xu, Y., Xu, C. X., and Sun, X. (**2004b**). "On the temporal domain of focus," Proceedings of the International Conference on Speech Prosody 2004, Nara, Japan.