Attenuation of vocal responses to pitch perturbations during Mandarin speech

Hanjun Liu

Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, People's Republic of China

Yi Xu

Department of Speech, Hearing and Phonetics Sciences, Division of Psychology and Language Sciences, University College London, London WC1N 1PF, United Kingdom

Charles R. Larson^{a)}

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

(Received 18 January 2008; revised 21 January 2009; accepted 22 January 2009)

The effect of stimulus timing on vocal responses to pitch-shifted feedback was investigated in different intonation patterns during Mandarin speech production. While speaking a four-word sentence consisting of the high-level tone, where the fundamental frequency (F_0) of the final word was either increased (question intonation) or slightly falling (statement intonation), pitch-shift stimuli (± 100 cents, 200 ms duration) were presented at three different times (160, 240, or 340 ms) after vocal onset. Results showed that in the question intonation, response magnitudes (16 cents) were significantly reduced for the 340 ms condition compared to the 160 (26 cents) or 240 (23 cents) ms conditions. No significant differences were found, however, as a function of stimulus timing in the statement intonation. These findings demonstrate that a planned change in F_0 can cause a modulation in the reflexive response to a perturbation in voice pitch feedback and that there is a critical time period during which the response mechanisms are most sensitive to the planning process. These findings suggest an approach for the study of mechanisms involved in the timing of successive words during speech. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3081523]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Jt, 43.70.Mn [AL]

Pages: 2299–2306

I. INTRODUCTION

In speech the role of voice fundamental frequency (F_0) is to convey various meanings in addition to those conveyed by consonants and vowels. Recent research on intonation has shown that F_0 can be used to encode two or more communicative meanings simultaneously (Eady and Cooper, 1986; Pierrehumbert and Hirschberg, 1990; Xu, 1999; Liu and Xu, 2005). It has been found that Mandarin speakers express lexical tone, focus, and question intonation by controlling F_0 in different ways (Xu, 1999; Liu and Xu, 2005). Tone is expressed by making F_0 approach a specific trajectory during the syllable, focus is encoded by expanding the pitch range of focused words and compressing the pitch range of post-focused words, and question is expressed by increasingly raising F_0 toward the end of the utterance. Similarly, simultaneous control of F_0 by lexical, focal, and intonational functions has been found in non-tonal languages such as English (Xu and Xu, 2005; Liu and Xu, 2007).

Important as it is to speech, little is known about neural mechanisms controlling voice F_0 . Recent attempts to model this system provide a theoretical rationale on how the nervous system might accomplish this task (Hain *et al.*, 2000). An important component of such models is the role of sen-

sory feedback in the control process. Experimental studies have demonstrated that if voice auditory feedback is missing, masked, or altered, voice F_0 control is diminished (Elliott and Niemoeller, 1970; Svirsky *et al.*, 1992; Mürbe *et al.*, 2002). The role of auditory feedback in voice control during speech development is particularly important, and deafness during language acquisition severely affects the quality of a child's articulation (Oller and Eilers, 1988). While the voice is not affected immediately if hearing loss occurs after mastering speech skills, there is a slow reduction in F_0 control (Angelocci *et al.*, 1964; Binnie *et al.*, 1982; Leder *et al.*, 1987; Waldstein, 1990; Lane *et al.*, 1997). It has also been shown that changes in kinesthetic feedback through anesthesia procedures reduce fine control of voice F_0 in pitch-masking tasks (Mallard *et al.*, 1978; Sundberg *et al.*, 1995). Recently, Larson *et al.* (2008) reported that voice responses to pitch perturbations were larger when the vocal fold mucosa was anesthetized than during normal kinesthesia, which could be due to the involvement of both kinesthesia and auditory feedback control of voice F_0 . In addition, it has been demonstrated that auditory feedback plays an important role in voice control in birds and mammals. For example, presenting electronically pitch-shifted voice feedback to horseshoe bat elicits compensatory changes in call frequency (Metzner, 1996), and the Lombard effect has been demonstrated in monkeys (Sinnott *et al.*, 1975). All of these studies

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

point to the importance of auditory feedback for on-line control of speech and vocal communication in general (Smotherman, 2007).

In addition to feedback models of voice F_0 control, it has also been suggested that speech and F_0 are controlled by “internal models” (Wolpert *et al.*, 1995; Perkell *et al.*, 1997; Kawato, 1999) that represent the relation between motor commands and their acoustic output. The development and maintenance of such internal models may depend on sensory feedback (Wolpert *et al.*, 1998; Jones and Munhall, 2000). That is, the neural system learns how to generate motor commands that can produce the desired acoustic output and to continuously fine-tune the internal representation of such motor-acoustic relations. Jones and Munhall (2000, 2002) showed that both English and Mandarin speakers compensated for pitch-shifted voice feedback to maintain their habitual pitch targets when their voice auditory pitch feedback was slowly shifted up or down without their awareness, indicating that auditory feedback is used in calibrating the internal model for the control of F_0 . When the pitch-shifted voice feedback was then suddenly removed, there was a negative aftereffect, which further supported the contention that the feedback had temporarily altered the internal model.

Others have also used the pitch-shift perturbation technique for the study of the role that auditory feedback plays in control of voice F_0 . When voice pitch feedback is unexpectedly either increased or decreased during vowel phonation or speech production, subjects compensate for the perturbation with voice F_0 responses that change in a direction opposite to that of the perturbation as if to nullify an unplanned deviation in F_0 (Elman, 1981; Kawahara, 1995; Burnett *et al.*, 1998; Hain *et al.*, 2000; Natke and Kalveram, 2001; Jones and Munhall, 2002). These compensatory responses have been interpreted as support for the idea that auditory feedback can be used online to help stabilize voice F_0 . Other studies have shown that voice F_0 responses to perturbations in voice pitch feedback may also be observed during speech (Donath *et al.*, 2002; Jones and Munhall, 2002; Xu *et al.*, 2004; Chen *et al.*, 2007). Moreover, such responses occur rapidly enough to correct for errors within the same syllable (Xu *et al.*, 2004) or subsequent syllables (Donath *et al.*, 2002; Xu *et al.*, 2004).

Recently, several studies of pitch-shifted voice feedback demonstrated that voice F_0 responses to pitch perturbations could be modulated according to the demands of the vocal task during either vowel phonation (Liu and Larson, 2007) or speech production (Natke *et al.*, 2003; Xu *et al.*, 2004). Task-dependent modulation of such responses was tested more directly in a study where English speakers either spoke a phrase or produced vowel phonations (Chen *et al.*, 2007). During speech production, vocal responses to downward pitch-shift stimuli presented before a rising F_0 contour were larger than to upward stimuli. Furthermore, overall responses to pitch perturbations during speech production were larger than those during vowel phonations. The larger responses to the downward stimuli during speech may indicate that the system recognized that the downward stimuli were in the wrong direction to the planned upward F_0 contour, and a corrective response was needed. Implicit in this assumption

is the idea that the F_0 level existing prior to a planned change in F_0 is known or can be predicted from the present F_0 level. If the actual F_0 level at the time of implementation of the change deviates from the predicted level, a response would be generated to correct for the error and to complete the planned change accurately. Thus the system controlling voice F_0 evidently takes into consideration the present F_0 level while planning for future changes in F_0 . If the system plans ahead for future changes in F_0 , there must be a time period during which this planning takes place.

The present study was designed to determine the boundaries of the temporal window during which planning for a subsequent change in F_0 takes place. Results from the study of Chen *et al.* (2007) suggest that the system is more sensitive to pitch-shift stimuli before a planned change in F_0 takes place. Therefore, measurement of response magnitudes to pitch-shift stimuli presented at different times before a planned change in F_0 may reveal the temporal boundaries of these periods of increased sensitivity to pitch-shift stimuli. In the present study, Mandarin subjects spoke a phrase that contained different intonation patterns while hearing perturbations in their voice pitch feedback. Pitch-shift stimuli were presented at three different times prior to a planned change in F_0 to test the hypothesis that response magnitude or latency would be modulated according to the timing of stimulus presentation in relation to the planned F_0 trajectory.

II. METHODS

A. Subjects

Ten native speakers of Mandarin (3 males and 7 females) between the ages of 24 and 30 were recruited. 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 and were paid for their participation.

B. Apparatus

Subjects were seated in a sound-treated room. They wore Sennheiser headphones with attached microphone (model HMD 280). The voice signal from the microphone was amplified with a Mackie mixer (model 1202), shifted in pitch with an Eventide Eclipse Harmonizer (H3000, SE), mixed with a 40 dB SPL pink masking noise (Goldline model PN2; spectral frequencies of 1–5000 Hz) through another Mackie mixer (model 1202-VLZ), and further amplified to 10 dB greater at the headphones than at the microphone with a Crown D75 amplifier and HP 350 dB attenuators. This 10 dB gain in feedback over the voice was used to reduce the chances that the subjects heard their non-shifted bone-conducted signal during vocalization. MIDI software (MAX/MSP, Version 4.5 by Cycling'74) on a laboratory computer was used to control the harmonizer and to generate control (TTL) pulses. A Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100) were used to make acoustic calibrations prior to testing any of the subjects. Subjects monitored their voice loudness from a Dorrough loudness monitor (model 40-A), providing a vi-

FIG. 1. Illustration of the timing of the stimuli relative to the F_0 contours for the question intonation (top) and statement intonation (bottom). The dashed line boxes indicate the timing and duration of the pitch-shifted stimuli.

sual feedback reference of their voice amplitude. Digitization of the voice output signal, feedback, and control pulses were performed at 10 kHz (12 bits), low-pass filtered at 5 kHz through POWERLAB (model ML880, AD Instruments), and recorded on a computer using CHART software (AD Instruments).

C. Procedures

Subjects were instructed, upon hearing one of two pre-recorded speech samples (male voice, ≈ 80 dB SPL) over the headphones, to repeat them with the same exact inflection pattern within 1 s. The Mandarin sentences consisted of four syllables, “mao1 mi1 mo1 ma1” (kitty touches mom), in which the numeral 1 represents the high-level tone. The pitch on the final syllable “...ma1” was raised in the question intonation and held constant in the statement intonation. The MIDI program, triggered by the onset of the subject’s voice, delivered a signal to the harmonizer that modulated the voice pitch feedback of the subject’s voice. The variability in the timing of the MIDI output from the onset of the pulse from the vocal detection circuit was about 25 ms. In each block of 60 trials, the model presented to the subjects on each trial had the same intonation pattern, and the stimulus timing was held constant at one of the three values. On each trial, the voice pitch feedback was either shifted up (increasing), down (decreasing), or not changed (control trials). The sequence of stimuli was randomized so that subjects could not predict which type of stimulus would occur on any given trial. Thus the subjects received 20 increases in pitch feedback, 20 decreases, and 20 control trials for each condition. Across six blocks (three stimulus timings by two intonation patterns) of 60 trials each, the stimulus magnitude was fixed at ± 100 cents (200 ms duration), but the timing was varied at 160 ms, 240 ms and 340 after vocal onset in the question and the statement intonations, as shown in Fig. 1. Table I lists the averaged timing values of pitch perturbations for three stimuli in the question and statement intonation. During any block of the trials, the stimulus timing was held constant at

TABLE I. Averaged timing values in milliseconds (SD) of stimuli relative to the onset of syllables 1–4 as a function of intonation, in which positive and negative values, respectively, indicate the presentation of pitch perturbation after and before the utterance of the syllable.

Stimulus (ms)	Intonation	Syllable 1	Syllable 2	Syllable 3	Syllable 4
160	Question	212 (26)	43 (31)	–150 (30)	–369 (38)
	Statement	204 (9)	13 (20)	–207 (28)	–438 (44)
240	Question	292 (25)	128 (22)	–69 (21)	–318 (85)
	Statement	282 (13)	87 (22)	–134 (37)	–381 (43)
340	Question	388 (17)	217 (20)	–39 (14)	–288 (24)
	Statement	387 (19)	196 (16)	–36 (18)	–279 (42)

one of the three delay times. Because of the natural variations in speaking rate, we attempted to keep speaking rate constant by having the subjects repeat a model, even though it was recognized that variation in timing would still occur. We also judged that attempting to measure the syllable durations for each subject would be impractical because here again, rates and syllable durations would vary. We therefore concluded that selecting a single set of delay times was the best compromise between experimental design and variations in production both within and across subjects.

For data analysis, the voice signal was processed in PRAAT (Boersma, 2001) 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 F_0 analog wave in IGOR PRO (Wavemetrics, Inc., Lake Oswego, OR). An operator manually marked the beginning and end points of the F_0 wave for each sentence according to the voice waveform, and then all the vocalizations in a block were time-normalized (linear interpolation) to reduce temporal variability and enhance the averaging procedures. An average of voice F_0 response was calculated by triggering the averaging computer on the TTL control pulse separately for each condition (phrase type, stimulus onset delay, and stimulus direction), summing all trials and dividing by N to produce an average F_0 response for that particular condition. A “difference wave” was then calculated by subtracting the average control wave from the average test waves derived from the increasing or decreasing pitch-shift stimulus. A point-by-point series of t -tests was run between all control and all test trials for a given condition and subject (Xu *et al.*, 2004). Valid responses were defined by significant t -tests ($p=0.02$) beginning at least 60 ms after the stimulus onset and lasting at least 50 ms. Response latencies were defined as the point where the p values of significant differences decreased below 0.02 and remained decreased for at least 50 ms. Response measurements were made during

TABLE II. Total number of “following” (FOL), “opposing” (OPP), and “non-response” (NR) across stimulus direction.

	Up	Down	Total
OPP	55	50	105
FOL	3	4	7
NR	2	6	8
Total	60	60	120

TABLE III. Total number of following (FOL), opposing (OPP), and non-response (NR) across stimulus timing.

	160 ms	240 ms	340 ms	Total
OPP	33	38	34	105
FOL	1	1	5	7
NR	6	1	1	8
Total	40	40	40	120

the 200 ms period of the stimulus because in the 340 ms condition, the intonated rise in F_0 often began near the end of the 200 ms making it difficult to measure the response. Response magnitudes were measured from the difference wave at the time indicated by the most significant p value within the above-defined measurement window. Statistical analyses were performed on response latency and magnitude using repeated-measures ANOVAs (SPSS, Version 16.0). Assumptions of compound symmetry and circularity for a repeated-measures ANOVA were met. For statistical analysis, non-responses were replaced by the mean value calculated from the measured data from other subjects for that condition.

III. RESULTS

From ten subjects across three stimulus timings, two stimulus directions, and two intonation patterns, there were 120 possible responses ($10 \times 3 \times 2 \times 2$). Tables II–IV list the numbers of opposing (the voice response change was in the opposite direction to the stimulus), “following” (response and stimulus changes in the same direction), and non-responses across stimulus direction, stimulus timing, and intonation pattern, respectively. 88% of the responses “opposed” the stimulus direction. Only 8 out of 120 responses did not meet our criteria of valid response (see Sec. II C) and were declared to be non-responses. The number of valid responses did not vary greatly across the experimental conditions.

Figures 2 and 3 illustrate average responses from representative subjects to pitch-shifted stimuli (thick black lines with error bars) superimposed on average control curves (thin black lines with error bars) across three stimulus timings in the question and statement intonations, respectively. Responses to the upward stimuli are illustrated on the left and downward stimuli on the right. The square brackets at the bottom of these figures indicate the time and direction of the stimulus. The major increase in traces in Fig. 2 coincides with the onset of the fourth syllable. The separation between the control (thin lines) and test (thick lines) wave that begins during the stimulus is the response to the perturbation. Compensatory vocal responses were produced across all experi-

TABLE IV. Total number of following (FOL), opposing (OPP), and non-response (NR) across intonation pattern.

	Question	Statement	Total
OPP	48	57	105
FOL	5	2	7
NR	7	1	8
Total	60	60	120

FIG. 2. Average F_0 contours for the test (thick lines) and control (thin lines) F_0 contours for stimulus timings of 160, 240, and 340 ms in the question pattern. Error bars represent the standard error of the mean for a single direction. Square brackets at the bottom indicate the time and the direction of the stimulus.

mental conditions in Figs. 2 and 3, except for the upward response for the 340 ms timing in the question intonation (bottom left in Fig. 2) that is a following response. In Fig. 2 (question intonation), it appears that response magnitudes for the responses in the 340 ms timing condition (right) are smaller than those for the 160 and 240 timing conditions. All the responses started before the rising of the F_0 contour of the last syllable. In Fig. 3 (statement intonation), there are no obvious differences in response magnitude across the three stimulus timing conditions.

Figures 4 and 5 show box plots of response magnitudes and latencies as a function of stimulus timing and direction in the question and statement intonations, respectively. Values of response magnitude and latency across stimulus tim-

FIG. 3. Average F_0 contours for the test (thick lines) and control (thin lines) F_0 contours for stimulus timings of 160, 240, and 340 ms in the statement pattern. Square brackets at the bottom indicate the time and the direction of the stimulus.

FIG. 4. Box plots illustrating the response magnitude as a function of stimulus timing and stimulus direction in the rising and the flat intonations. Box plot definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, and 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).

ing and intonation pattern are shown in Tables V and VI. Three-way repeated-measures ANOVAs (stimulus direction, stimulus timing, and intonation pattern) performed on the response magnitude revealed significant main effects for stimulus timing [$F(2,18)=6.979, p=0.006$] but not for stimulus direction [$F(1,9)=2.083, p=0.183$] or intonation pattern [$F(1,9)=0.008, p=0.993$]. A significant interaction effect was found between intonation pattern and stimulus timing [$F(2,18)=6.826, p=0.006$]. In order to determine the effect of stimulus timing on vocal responses in the different intonations, two-way repeated-measures ANOVAs were performed on the response magnitude and latencies for the rising and the flat intonation patterns. The results indicated significant main effects across stimulus timing in the rising [$F(2,18)=9.913, p=0.001$] but not in the flat pattern [$F(2,18)=1.276, p=0.303$]. No significant main effects were found for stimulus direction in the question [$F(1,9)=0.286, p=0.625$] or the statement condition [$F(1,9)=7.028, p=0.250$], and there were no interaction effects between stimulus direction and stimulus timing for the intonation patterns [question: $F(2,18)=0.390, p=0.768$; statement: $F(2,18)=3.236, p=0.180$]. *Post hoc Bonferroni* tests in the rising intonation indicated that the 160 ms (26 ± 7 cents) and 240 ms (23 ± 9 cents) conditions produced significantly larger response magnitudes than the 340 ms (16 ± 6 cents) condition ($p=0.001$ and $p=0.014$, respectively). In addition, three-way repeated-measures ANOVAs performed on the response latency revealed no significant main effects across all conditions.

FIG. 5. Box plots illustrating the response latency as a function of stimulus timing and stimulus direction in the rising and the flat intonations.

IV. DISCUSSION

The purpose of the present study was to investigate the effect of timing of pitch-shifted voice feedback on the vocal response across two different intonation patterns during meaningful Mandarin speech. The findings indicated significantly larger response magnitudes for the 160 and 240 ms stimulus onset conditions than for the 340 ms condition in the question intonation, while no significant differences existed between 160 and 240 ms timing conditions. Furthermore, there was no effect of stimulus timing on the responses in the statement intonation.

It is now well known that perturbations in voice pitch feedback evoke compensatory changes in voice F_0 during sustained vowels, singing and speaking conditions (Elman, 1981; Kawahara, 1995; Burnett *et al.*, 1998; Hain *et al.*, 2000; Natke and Kalveram, 2001; Donath *et al.*, 2002; Jones and Munhall, 2002; Xu *et al.*, 2004; Chen *et al.*, 2007). The mechanisms of these responses are unknown but necessarily

TABLE V. Average response magnitudes in cents (SD) as a function stimulus timing and intonation pattern, in which values were collapsed across the stimulus direction.

	160 ms	240 ms	340 ms	Mean
Rising	26 (7)	23 (9)	16 (6)	21 (9)
Flat	23 (9)	22 (8)	21 (9)	22 (9)
Mean	24 (8)	23 (9)	19 (8)	22 (9)

TABLE VI. Average response latencies in milliseconds (SD) as a function stimulus timing and intonation pattern, in which values were collapsed across the stimulus direction.

	160 ms	240 ms	340 ms	Mean
Question	112 (43)	98 (30)	100 (45)	104 (39)
Statement	125 (60)	114 (56)	102 (42)	113 (53)
Mean	118 (52)	106 (45)	101 (43)	109 (47)

involve detection of a change in voice pitch feedback, comparison of the feedback with an internal referent of the desired F_0 , activation of neurons related to motor output, transmission delays in the central and peripheral nervous system, and contraction (or relaxation) of laryngeal muscles leading to a change in voice F_0 . Question intonation in Mandarin involves a large increase in F_0 at the utterance-final position (Ho, 1977; Lin, 2004; Liu and Xu, 2005), although smaller F_0 increases also occur before the final syllable (Yuan *et al.*, 2002; Liu and Xu, 2005; Ni *et al.*, 2006). For statements, it has been found that there is an F_0 decline over the course of an utterance in Mandarin (Xu, 1999; Shih, 2000), but the declination is very slight if the utterance consists of only the high-level tone (Xu, 1999). These patterns are also found in the utterances produced by our subjects. Large F_0 differences occurred at the utterance-final position between questions (288 Hz) and statements (241 Hz), [$F(1,9)=51.593, p=0.000$] for this mix of male and female subjects. Smaller F_0 differences also occurred in the first syllable, 257 Hz in questions and 240 Hz in statements [$F(1,9)=5.753, p=0.040$].

When a speaker plans on increasing voice F_0 at the end of an utterance in a question, as is the case in the present study, there must be a planning process that is aware of the current F_0 level and the correct time to contract muscles to change F_0 . The observations of the present study suggest that there is a time-dependent interaction between responses to pitch perturbations and mechanisms involved in planning a question intonation in Mandarin speech. The interaction is time-dependent because only responses to the 340 ms stimuli, not responses to stimuli that occurred earlier, were affected. The interaction is related to the planning of the increase in F_0 because the reduction in response magnitude only occurred with the question intonation condition, not the statement intonation condition.

Further observations suggest the dimensions of the time window during which planning for the F_0 contour may take place. With the 340 ms stimulus onset time, the response began about 100 ms following the stimulus onset and about 100 ms before the rising F_0 contour of the fourth syllable. The decrement in response magnitude was not observed for the 160 or 240 ms conditions with the rising intonation (question). Therefore, the reduction in the response is related to the rising intonation on the last syllable. Moreover, the time period when the interaction between the pitch-shift stimulus and the decrement in response magnitude occurred must be between the stimulus onset, 200 ms before the rising intonation, and the beginning of the response, 100 ms before the rising intonation. It is therefore suggested that there is a time window of 100 ms duration (between 200 and 100 ms

before a planned increase in F_0 contour) during which auditory feedback may interact with the central planning mechanisms.

It is next instructive to consider the reasons for this interaction. One explanation for the reduction in response magnitudes observed for the 340 ms condition with the rising F_0 contour may relate to the pattern of voice F_0 regulation in question intonation. When a sentence is spoken as a question with either final or neutral focus, F_0 starts to rise exponentially from the start of the last word (Liu and Xu, 2005). In the present study, the final word is the syllable /ma/ for “mother.” In other words, the F_0 of /ma/ is controlled by a lexical function as well as an intonation function. Such dual control presumably requires greater amounts of neural resources than when only a single function is involved. This control may have depleted some of the neural resources needed to monitor sensory, including auditory feedback.

Another possible explanation for these results is that the auditory perturbation interacted with mechanisms that were involved in elevating F_0 at the end of the sentence. One type of interaction may be that the mechanisms responsible for the voluntary increase in F_0 caused a reduction in sensitivity to the perturbation so that the response to the perturbation would not interfere with the planned increase in F_0 . It is also possible that there was active attenuation of the responding mechanisms, again possibly to avoid interference with the planned rise in F_0 . Regardless of the exact mechanisms and the reasons for the response reduction, the results suggest that there is a time window (100–200 ms) before a rise in F_0 during which a response to an auditory perturbation may interact with speech planning mechanisms.

Another interpretation of the data relates to the method of measurement of the responses. Due to the fact that for the question intonation condition, there was a rising intonation on the fourth syllable and the need to make response measurement conditions consistent across all conditions, it was necessary to limit the time window for measurement of responses in all conditions to the 200 ms stimulus window. If it had been possible to extend the measurement window out further in time, response magnitudes may have been the same in all timing conditions. Nevertheless, the fact that the responses in the 340 ms condition were smaller than those for the 160 and 240 ms conditions within a 200 ms time period following the stimulus indicates that response-generating mechanisms behaved differently in the 340 ms condition compared with the other timing conditions.

A. Comparison with other studies

The effect of stimulus timing on the response during the production of bi-tonal Mandarin phrase was also investigated by Xu *et al.* (2004). The disyllabic nonsense phrases /ma ma/ with different tonal patterns (high-high, high-rising, and high-falling) were tested with stimuli of 100 and 250 ms delays after the onset of the voice. The results indicated no effect of stimulus timing on the response magnitude or the latency. This finding is in contrast to the present study and may be explained by the fact that only tonal changes were manipulated in the study of Xu *et al.*, without intonation

changes, which may require a different response mechanism during the dynamic control of voice F_0 . Another explanation is that a nonsense phrase was used in the study of Xu *et al.* (meaningful words in a nonsense phrase), and meaningful speech was tested in the present study. In the present study, however, data in the statement intonation indicated no effect of stimulus timing on the response, which is comparable with the findings of Xu *et al.* in the high-high (/ma ma/) condition.

In comparing this study with previous pitch shift experiments, the variability in speech style (meaningful speech vs nonsense speech) may contribute to the variability in response magnitudes and latencies. For example, during nonsense speech, vocal response magnitudes of 47 cents (Natke *et al.*, 2003) were reported for 100 cents stimuli, and magnitudes of 50 cents (static tone) and 85 cents (dynamic tone) were reported for 200 cents stimuli (Xu *et al.*, 2004). These values are larger than the overall response magnitudes during meaningful speech of 22 cents in the present study (see Table V). The smaller magnitude responses reported for meaningful speech compared to nonsense speech in these studies may be partially explained by findings from the present study that intonation pattern and stimulus timing can modulate response magnitudes. It should come as no surprise, therefore, if responses to pitch-shifted feedback during some specific types of speech production conditions may be smaller than those in nonsense speech.

One problem with the interpretation of the data in the present study comes from a comparison of the present results with those reported by Chen *et al.* (2007). In that study, responses to 200 cents downward stimuli that were presented prior to a rise in F_0 during English speech were larger than those to upward stimuli. In the present study, there was no difference in magnitude of responses between upward and downward stimuli in any of the timing conditions. The 340 ms condition was most similar to that tested in the study of Chen *et al.*, and it is surprising that in the present study responses to the downward stimuli were not larger than those to the upward stimuli, as they were in the study of Chen *et al.* Two factors may explain this. First, although the rise in F_0 at the end of the phrase was a supra-segmental adjustment in both studies, the Mandarin language background of the subjects in the present study may have predisposed them to respond differently than the English speakers in the study of Chen *et al.* A recent finding has shown that speakers of English produce a much larger final rise in question intonation than do speakers of Mandarin (Liu and Xu, 2007). Second, the effect of the increase in response magnitude in the study of Chen *et al.* was only seen with 200 cents stimuli, not 100 cents stimuli as used in the present study. That is, the stimulus magnitude in the present study may have been too small to evoke the directional difference in the auditory system, causing no significant differences between the responses as a function of stimulus direction.

With respect to response latency, mean values of 211 ms (Jones and Munhall, 2002), 165 ms (Xu *et al.*, 2004), and 150 ms (Natke *et al.*, 2003) were reported for nonsense speech, while 122 ms (Chen *et al.*, 2007) and 109 ms (see Table VI) were observed in the present study during mean-

ingful speech. One explanation for these differences may be that they are attributable to methodologies. Jones and Munhall used a value of 2 standard deviations (SDs) above or below the pre-stimulus mean to define the onset of a response. Natke *et al.* used the Castellan change-point test to define the “change point” in F_0 contours following a stimulus. The methods for measuring response latency in the present study are identical to those used in the studies of Xu *et al.* (2004) and Chen *et al.* (2007) and are hence more comparable to those studies.

Other than the methodological differences, the latency differences suggest that responses to pitch-shifted voice feedback may be faster during meaningful speech than during nonsense speech. This observation may be explained by considering that the target of natural speech is to communicate with listeners. Even though response magnitudes during meaningful speech may be smaller than those during nonsense speech in some conditions, it may be more important for communication effectiveness for subjects to correct for errors faster during meaningful speech than during nonsense speech, resulting in the shorter latencies during meaningful speech than during nonsense phrases.

V. CONCLUSION

In the present study, the effect of the stimulus timing on the pitch-shift reflex during Mandarin speech was investigated in different intonation patterns. The results revealed that stimulus timing is an important variable that can affect the magnitude of vocal responses to pitch perturbations when they occur in a question intonation. Responses to pitch perturbations that were presented just prior to a planned pitch increase were reduced in magnitude compared to stimuli that were presented earlier or to those that were presented when there was no planned F_0 change. The reduction in response magnitude to stimuli that are presented just before a planned increase in voice F_0 demonstrates a contrast to previous reports that showed an increase in response magnitude in some speaking conditions. In combination, these contrasting effects demonstrate that the audio-vocal system can regulate the vocal response depending on the specific task. The factors that appear to be responsible for these differing effects are the timing of the stimulus and whether or not a subsequent change in F_0 is planned. Future studies directed at defining the time window in which stimuli can lead to changes in response magnitude may help us to understand the timing of neural mechanisms involved in the control of temporal features of speech.

ACKNOWLEDGMENT

This work was supported by NIH Grant No. 1R01DC006243. We thank Mr. Chun Liang Chan for programming assistance and Ms. Dilpreet Kaur Minhas for her assistance in data analysis.

Angelocci, A. A., Kopp, G. A., and Holbrook, A. (1964). “The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys,” *J. Speech Hear. Disord.* **29**, 156–170.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W. (1982). “Phonetic disintegration in a five-year-old following sudden hearing loss,” *J. Speech*

- Hear. Disord. **47**, 181–189.
- Boersma, P. (2001). PRAAT, a system for doing phonetics by computer (Glot International).
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F_0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice 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.
- 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.
- Elliott, L., and Niemoeller, A. (1970). "The role of hearing in controlling voice fundamental frequency," Int. Audiol. **9**, 47–52.
- Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.
- 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.
- Ho, A. T. (1977). "Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative," Phonetica **34**, 446–457.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F_0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.
- 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," in 'Computational Auditory Scene Analysis' and 'International Joint Conference on Artificial' Intelligence, Montreal.
- Kawato, M. (1999). "Internal models for motor control and trajectory planning," Curr. Opin. Neurobiol. **9**, 718–727.
- Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., and Manzella, J. (1997). "Changes in sound pressure and fundamental frequency contours following changes in hearing status," J. Acoust. Soc. Am. **101**, 2244–2252.
- Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (2008). "Interactions between auditory and somatosensory feedback for voice F_0 control," Exp. Brain Res. **187**, 613–621.
- Leder, S. B., Spitzer, J. B., and Kirchner, J. C. (1987). "Speaking fundamental frequency of postlingually profoundly deaf adult men," Ann. Otol. Rhinol. Laryngol. **96**, 322–324.
- Lin, M. (2004). "On production and perception of boundary tone in Chinese intonation," in *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (Chinese Academy of Social Sciences, Beijing), pp. 125–129.
- Liu, F., and Xu, Y. (2005). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," Phonetica **62**, 70–87.
- Liu, F., and Xu, Y. (2007). "Question intonation as affected by word stress and focus in English," in *Proceedings of the 16th International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry (International Congress of Phonetic Sciences, Saarbrücken), pp. 1189–1192.
- Liu, H., and Larson, C. R. (2007). "Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex," J. Acoust. Soc. Am. **122**, 3671–3677.
- Mallard, A. R., Ringel, R. L., and Horii, Y. (1978). "Sensory contributions to control of fundamental frequency of phonation," Folia Phoniatr. Logop. **30**, 199–213.
- Metzner, W. (1996). "Anatomical basis for audio-vocal integration in echolocating horseshoe bats," J. Comp. Neurol. **368**, 252–269.
- Mürbe, D., Pabst, F., Hofmann, G., and Sundberg, J. (2002). "Significance of auditory and kinesthetic feedback to singers' pitch control," J. Voice **16**, 44–51.
- 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.
- Ni, J., Kawai, H., and Hirose, K. (2006). "Constrained tone transformation technique for separation and combination of Mandarin tone and intonation," J. Acoust. Soc. Am. **119**, 1764–1782.
- Oller, D., and Eilers, R. (1988). "The role of audition in infant babbling," Child Dev. **59**, 441–449.
- Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (1997). "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**, 227–249.
- Pierrehumbert, J., and Hirschberg, J. (1990). in *Intentions in Communication*, edited by P. R. Cohen, J. Morgan, and M. E. Pollack (Massachusetts Institute of Technology Press, Cambridge, MA), pp. 271–311.
- Shih, C. (2000). in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Heidelberg), pp. 243–268.
- Sinnott, J. M., Stebbins, W. C., and Moody, D. B. (1975). "Regulation of voice amplitude by the monkey," J. Acoust. Soc. Am. **58**, 412–414.
- Smotherman, M. S. (2007). "Sensory feedback control of mammalian vocalizations," Behav. Brain Res. **182**, 315–326.
- Sundberg, J., Iwarsson, J., and Billstrom, A. H. (1995). "Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers," J. Voice **9**, 20–26.
- Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (1992). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.
- Waldstein, R. (1990). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). "An internal model for sensorimotor integration," Science **269**, 1880–1882.
- Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). "Internal models in the cerebellum," Trends Cogn. Sci. **2**, 338–347.
- Xu, Y. (1999). "Effects of tone and focus on the formation and alignment of F_0 contours," J. Phonetics **27**, 55–105.
- 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.
- Xu, Y., and Xu, C. X. (2005). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.
- Yuan, J., Shih, C., and Kochanski, G. P. (2002). "Comparison of declarative and interrogative intonation in Chinese," in *Proceedings of the First International Conference on Speech Prosody* (International Conference on Speech Prosody, Aix-en-Provence, France), pp. 711–714.