Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech

Hanjun Liu and Qianru Zhang
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 7 November 2006; revised 25 July 2007; accepted 30 July 2007)

Previous studies have demonstrated that perturbations in voice pitch or loudness feedback lead to compensatory changes in voice F_0 or amplitude during production of sustained vowels. Responses to pitch-shifted auditory feedback have also been observed during English and Mandarin speech. The present study investigated whether Mandarin speakers would respond to amplitude-shifted feedback during meaningful speech production. Native speakers of Mandarin produced two-syllable utterances with focus on the first syllable, the second syllable, or none of the syllables, as prompted by corresponding questions. Their acoustic speech signal was fed back to them with loudness shifted by ±3 dB for 200 ms durations. The responses to the feedback perturbations had mean latencies of approximately 142 ms and magnitudes of approximately 0.86 dB. Response magnitudes were greater and latencies were longer when emphasis was placed on the first syllable than when there was no emphasis. Since amplitude is not known for being highly effective in encoding linguistic contrasts, the fact that subjects reacted to amplitude perturbation just as fast as they reacted to F_0 perturbations in previous studies provides clear evidence that a highly automatic feedback mechanism is active in controlling both F_0 and amplitude of speech production.
 [DOI: 10.1121/1.2773955]

PACS number(s): 43.72.Dv, 43.70.Mn, 43.70.Jt, 43.70.Gr [AL] Pages: 2405–2412

I. INTRODUCTION

There have been several studies of the mechanisms that control voice intensity. Considering peripheral mechanisms, subglottal air pressure, air flow, glottal impedance, voice fundamental frequency (F_0) and vocal tract impedance all interact to affect vocal intensity (Isshiki, 1964; Koyama *et al.*, 1969; Titze and Sundberg, 1992). The variations in vocal intensity that occur during speech are the result of interactions between the pressure-relaxation forces of the respiratory system, and respiratory and laryngeal muscle contractions (Draper *et al.*, 1959; Hirano and Ohala, 1969; Ladefoged and Loeb, 2002). Neural mechanisms of voice intensity control are less well understood. Lombard was the first to demonstrate the importance of auditory feedback on the control of intensity (see Lane and Tranel, 1971). It was found that the presence of environmental noise affected voice intensity where speakers raised their vocal intensity to make themselves heard over the noise level. Similarly, the phenomenon of side-tone amplification demonstrated that if a speaker's voice is amplified above a normal level, the speaker will reduce his or her intensity; if the feedback level of a person's voice is reduced, speakers will raise their intensity (Chang-Yit *et al.*, 1975; Garber *et al.*, 1976; Lane and Tranel, 1971; Siegel and Pick Jr., 1974).

As important as these studies are, the research paradigm used in them raises questions on their interpretation. In the typical paradigm, a person is instructed to read a passage in the presence of noise, side-tone amplification, or both side-tone amplification and noise. Generally these auditory feedback variables are present for the duration of the speaking task, and the intensity adjustments made by the speaker are considered to be automatic. In this sense, the word automatic does not necessarily mean reflexive, but rather an adjustment a speaker would naturally make to increase the effectiveness in communicating with others. Because of the reliability of the side-tone amplification effect, it was suggested that it may reflect a "fundamental characteristic of speech regulation" (Chang-Yit *et al.*, 1975, p. 324) to maximize the communicative effectiveness by increasing the signal to noise ratio. Despite the importance of regulating one's voice in an attempt to overcome noise or distance, it is not clear from the above studies whether subjects monitor voice feedback and make corrections online in case the production does not match that which was intended.

Two recent studies have demonstrated that auditory feedback seems to play a role in the online control of voice intensity during vowel production. Heinks-Maldonado and

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

Houde (2005) and Bauer *et al.* (2006) demonstrated that during vowel productions, speakers would respond to brief (e.g., 200 ms) perturbations in voice loudness feedback by making changes in their voice amplitude. The latencies of the responses, with a mean value between 150 and 300 ms, coupled with the fact that the direction of the responses is generally opposite to the stimulus direction, regardless of whether the stimulus is an increase or decrease in loudness feedback, suggests these responses are reflexive in nature. These studies are similar to those that have been conducted to study the effects of voice pitch feedback perturbations on voice F_0 control, which are also thought to be reflexive in nature (Bauer and Larson, 2003; Burnett *et al.*, 1998; Hain *et al.*, 2000). Moreover, the fact that responses to pitch perturbations are observed in speech (Donath *et al.*, 2002; Jones and Munhall, 2002; Natke *et al.*, 2003), and increase in magnitude in some speaking conditions (Chen *et al.*, 2007; Xu *et al.*, 2004), raises the possibility that voice amplitude responses to perturbations in loudness feedback may also be present in speech. A third line of research demonstrating the importance of auditory feedback in speech comes from studies showing long term adaptation to changes in formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006).

However, the role of amplitude in speech is unclear, and whether or not amplitude is controlled like F_0 may depend on the importance of amplitude control during speech. While there is evidence that overall amplitude is actively controlled by speakers and that auditory feedback plays a role in this process (Bond and Moore, 1994; Chang-Yit *et al.*, 1975; Dreher and O'Neill, 1958; Garber *et al.*, 1976; Lane *et al.*, 1995; Lane *et al.*, 1997; Leder *et al.*, 1987; Siegel and Kennard, 1984; Svirsky *et al.*, 1992; Van Summers *et al.*, 1988), there have also been findings suggesting that, unlike F_0, amplitude is not highly effective in conveying communicative functions such as lexical stress, or focus (Fry, 1958; Turk and Sawusch, 1996). A recent analysis of natural speech databases, however, has found amplitude to be correlated with the perception of prominence (Kochanski *et al.*, 2005). Regardless of its actual function, it is an open question as to whether amplitude is rapidly adjusted in response to perturbations in voice loudness during the production of meaningful speech. If voice amplitude is adjusted online in response to perturbations in auditory feedback, it would suggest that amplitude control during speech is important for the expression of linguistic functions such as focus or word stress. The purpose of the present study was to investigate whether speakers compensate for perturbations in voice loudness feedback during speech.

The Mandarin speech stimuli that were investigated were designed to differ in their focus patterns. Focus is discourse-motivated emphasis, and is known to be accompanied by expansion of pitch range in the focused word and suppression of pitch range in the postfocus words (Xu, 1999; Xu and Xu, 2005). The pitch range expansion in a final focus is much smaller than in an earlier focus, which makes its percept less salient (Liu and Xu, 2005). Although there have not been systematic data, the pitch range changes due to focus may also be accompanied by amplitude variations. It is also possible that such amplitude variations are actively controlled. If they are actively controlled, and if auditory feedback plays a role in this control process, then responses to loudness-shifted voice feedback during a focused syllable should be greater than if the syllable is not focused.

II. METHODS

A. Subjects

Ten native speakers of Mandarin (six males and four females; ages 19–30), most of whom were students at Northwestern University, served as subjects. All subjects reported normal hearing, and none reported a history of neurological or communication disorders. All signed informed consent approved by the Northwestern Institutional Review Board and were paid for their participation.

B. Apparatus

Subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (model HMD 280) throughout the testing. They were asked to speak aloud the experimental stimuli 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. This monitor provided the subjects with visual feedback on their voice amplitude and helped them to maintain a relatively constant level throughout the testing. The rapid changes in the visual display, which coincided with changes in voice amplitude, were too fast for the subjects to respond to and thus do not affect the results. This feedback merely helped the subjects to maintain a relatively constant amplitude level throughout the testing. The vocal signal from the microphone was amplified with a Mackie mixer (model 1202), processed for loudness shifting with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise with a Mackie mixer (model 1202-VZL), further amplified with Crown D75 amplifier and HP 350 dB attenuators at 80 dB SPL, and sent back to the headphones. The harmonizer was controlled with MIDI software (Max/MSP v.4.1 by Cycling 74) from a laboratory computer. Acoustic calibrations were made with a Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100). There was a gain of 10 dB SPL between the subject's voice amplitude, measured 2.5 cm from the mouth, and the feedback loudness measured at the input to the ear canal. The voice output signal, feedback and control pulses (TTL) were digitized at 10 kHz, low-pass filtered at 5 kHz and recorded on a laboratory computer utilizing Chart software (AD Instruments). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR).

C. Procedures

The disyllabic sequence produced by the subjects was /ba1 ma1/, meaning "the eighth aunt." This phrase was produced in response to three different questions the subject would hear over the headphones, as shown in Table I. Each of the three questions required the subjects to produce the

TABLE I. Disyllabic list of questions and responses, where numerals 1, 2, 4 represent the High, Rising, and Falling tones, and the underscored syllable is focused.

Question	Response
/shui?/ "Who?"	/ba1 ma1/ (No Focus)
/ba4 ma1?/ "Father and mother?"	/<u>ba1</u> ma1/ (1st Focus)
/ba1 yi2/? "The eighth aunt in mother's family?"	/ba1 <u>ma1</u>/ (2nd Focus)

phrase with one of three different focus patterns: first syllable (1st Focus), second syllable (2nd Focus) or neither syllable (No Focus) (Xu, 1999).

Each subject completed a total of 180 trials that were divided into three groups of 60 trials with intervening rest periods. In each group of trials, 60 questions requiring 60 responses were presented. The three questions were randomly distributed among the group of 60. During the production of a group of 60 trials, voice loudness feedback was either increased, decreased or not changed (control trials). These stimuli were presented randomly so that the subjects could not predict which type of stimulus, or control (no stimulus), would occur on each trial. Thus, a subject received 20 increases in loudness feedback, 20 decreases and 20 control trials for each of the three questions. The stimuli were ±3 dB SPL perturbations (200 ms duration) in voice loudness feedback beginning 160 ms after vocal onset. Figure 1 shows example voice amplitude contours during production of the three phrases (solid lines) superimposed on the perturbed loudness feedback signal (dashed lines). In the first example there was a decrease in feedback loudness, in the middle example there was no change in feedback loudness (control) and in the 3rd example there was an increase in feedback loudness. During the recording, three TTL pulses were recorded that indicated the type of loudness shift stimulus (upward, downward, or control) that was presented. During data acquisition, the experimenter listened to the recorded vocal responses to insure that the subject responded with the correct response to each question. Incorrect responses were discarded from further analysis, and extra trials were run to insure that 180 total trials with the correct responses were obtained for each subject.

D. Data analysis

Digitized signals were analyzed by converting the voice signal to a root-mean-square (rms) voltage signal calculated using a 50 ms sliding window

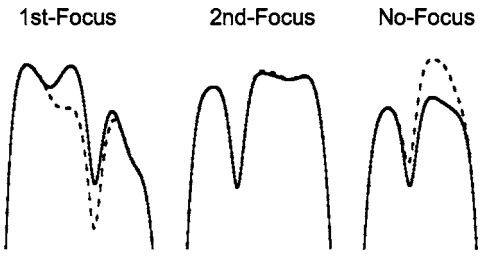

FIG. 1. Voice amplitude contours during production of the three phrases, 1st Focus, 2nd Focus and No Focus. Solid lines represent voice amplitude and dashed lines represent voice loudness feedback. These examples are for single productions. Time and magnitude scales are in relative units.

$$\mathrm{rms}(x) = \sqrt{\frac{1}{N}\sum_{n-25}^{n+25} x^2}, \tag{1}$$

where x is the value of each data point, and N is the total number of data points. Voice rms voltage measures were then converted to dB SPL to reflect actual voice amplitude using the following formula:

$$\mathrm{Voice(dB)} = 20 \times \log(\mathrm{rms}(x)/c) + 70, \tag{2}$$

where c equals 0.228, which is the rms voltage corresponding to a vocal level of 70 dB SPL that was obtained through calibration procedures. This wave and the TTL pulses were displayed on a computer screen. An operator marked the onset and offset of each vocalization (phrase) according to the voice wave form signal. The durations of all vocal signals along with the accompanying TTL signal representing timing and direction of the stimulus were then time normalized. Time normalization (linear interpolation) was done to reduce temporal variations in the speech signals and thereby reduce variability in the subsequent averaged trials. For each subject, an ensemble average of each set of test and control trials was generated for each speech and stimulus direction condition by triggering the averaging program at the onset of each of the TTL pulses.

A point-by-point series of t tests were then conducted between all the test and respective control trials for a given condition (Chen *et al.*, 2007; Xu *et al.*, 2004). The result of the t test was a set of "p" values representing the significance of the t test for each time point of the set of control and test waves. Wherever the "p wave" decreased below a value of 0.02 following the stimulus onset with a delay of at least 60 ms and remained low for at least 50 ms, the crossing point was designated as the onset of a response (latency). The temporal constraints imposed by these criteria guarded against spurious significant differences (Chen *et al.*, 2007). The point where the p wave increased to a value greater than 0.02 was defined as the response termination. A "difference" wave was then calculated by subtracting the average control wave from the averaged increasing and decreasing loudness stimulus test waves for each subject and each condition. Thus for each focus and perturbation condition there was a "difference wave" representing the response to an upward shift in loudness feedback compared with the control condition, and one representing the response to a downward shift in loudness feedback compared with the control condition. Using the times noted in the analysis of the p wave crossing the value of 0.02, a program then measured the peak (or trough) magnitude of the difference wave. If responses failed to reach significance within the above-described temporal parameters, they were designated as nonresponses. The time stamps of the p wave crossings along with the magnitude measures of the difference waves were tested for significance with a repeated-measures analysis of variance (ANOVA) (SPSS, v. 11.0). For statistical analysis, nonresponses were replaced by the mean value calculated from the measured data from other subjects for that condition. This procedure allowed us to use a repeated-measures design. Assumptions

TABLE II. Total number of "following" (FOL), "opposing" (OPP), and "nonresponse" (NR) across two stimulus directions.

	Up	Down	Total
FOL	11	9	20
OPP	14	16	30
NR	5	5	10
Total	30	30	60

of a normal distribution, homogeneity of variance, compound symmetry and circularity for a repeated measures ANOVA were met.

III. RESULTS

From ten subjects across the three experimental conditions and two stimulus directions (10*3*2), there were 60 possible responses. Tables II and III list the numbers of opposing, "following" and nonresponses across the two stimulus directions and three intonation patterns. Approximately 33.3% of the responses "followed" the stimulus direction. About 16.7% of the responses did not meet our criteria of validity and were declared to be nonresponses. 50% of responses opposed the direction of the loudness-shift stimulus. The percentage of valid responses did not vary greatly across the three experimental conditions. Also the numbers of valid responses by stimulus direction did not differ greatly (14 Up and 16 Down).

Figure 2 shows examples of the average voice amplitude contours for each of the three questions and stimulus direction. In each plot, the responses to loudness shift stimuli (thick lines) are shown along with the average control curves (thin lines). At the bottom of the plots, the square brackets represent the timing and direction of the stimulus. In all the plots, the first syllable is clearly separated from the second syllable by a drop in SPL. Due to the differences in the emphasis patterns, the duration of syllables /ba/ and /ma/ changed. The first syllable is longer when emphasis is on the first syllable than when it is on the second syllable or when there is no emphasis.

In Fig. 2 the drop in SPL associated with the transition to the second syllable occurred right after the stimulus onset. The stimuli persist into the second syllable, and the response, indicated by the divergence between the test and control contours, can be easily seen during the second syllable. The amount of the separation reflects the response magnitude. Arrows indicate the time at which the response magnitude was measured. Responses to increasing loudness stimuli are shown on the left and decreasing loudness stimuli on the right. In Fig. 2, the response to the increasing stimulus in the

TABLE III. Total number of "following" (FOL), "opposing" (OPP), and "nonresponse" (NR) across three phrase types.

	1st Focus	2nd Focus	No Focus	Total
FOL	6	7	7	20
OPP	9	11	10	30
NR	5	2	3	10
Total	20	20	20	60

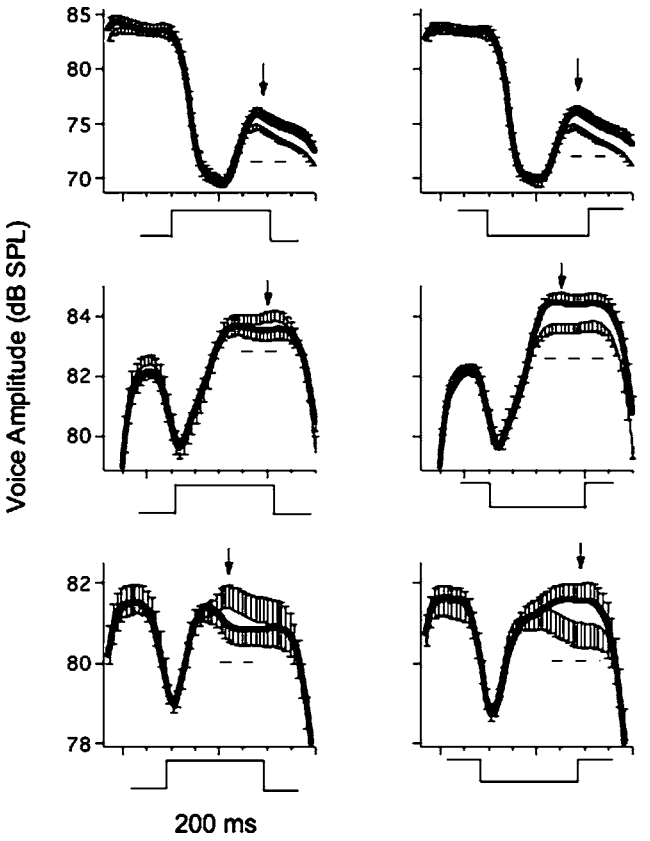

FIG. 2. Control (thin black line) and test average waves (thick black line) during 1st- Focus, 2nd-Focus, and No-Focus sequences (from top to bottom). Contours cover the entire duration of the utterances. The vertical arrow indicates time where the response magnitude was measured. Error bars represent the standard error of the mean for a single direction. Curves at the bottom indicate the time and the direction of the stimulus. Horizontal dashed lines indicate time period where the two waves differed significantly.

1st-Focus condition "follows" the direction of the stimulus. All other responses in this figure oppose the direction of the stimulus and are compensatory in nature.

Figures 3 and 4 show box plots of the response magnitude and latency across three phrase types and two stimulus directions, respectively. Values of response magnitudes and latencies are shown in Tables IV and V. Two-factor repeated-measures ANOVAs were performed on measures of response magnitude and latency across the factors phrase type and stimulus direction. For statistical analysis, measures for both the compensating and following responses were grouped together because there were only 50 total responses, and there was a relatively large number of following responses compared to previous studies (Chen *et al.*, 2007; Xu *et al.*, 2004). Moreover, there were no statistically significant differences in the measures of response magnitude ($F(1,49)=1.743$, $p=0.191$) or latency ($F(1,49)=0.964$, $p=0.330$) between the groups. For the response magnitude, a significant main effect was found for phrase type ($F(2,18)=3.923$, $p=0.039$, Fig. 3 and Table IV) but not for stimulus direction ($F(1,9)=0.991$, $p=0.345$; Fig. 4 and Table V). A post hoc test indicated that the responses in the 1st-Focus phrase were significantly larger than those in the 2nd-Focus phrase ($p=0.032$; post hoc Bonferroni; Table IV). Response magnitudes for the 2nd Fo-

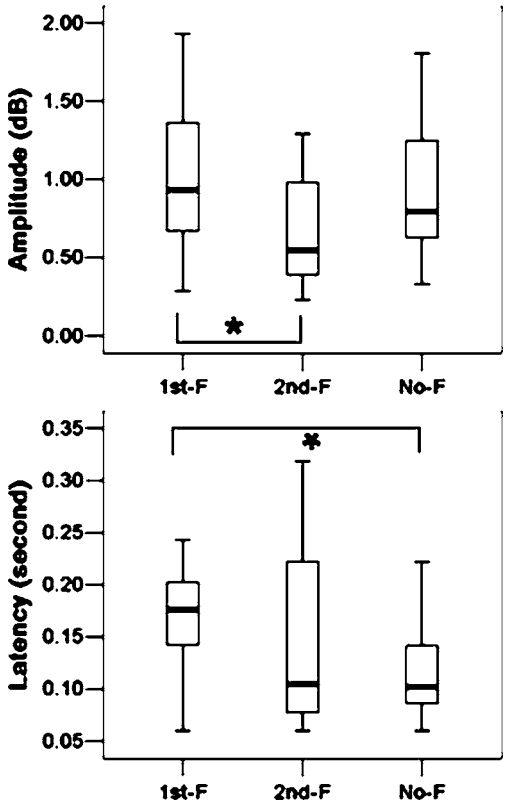

FIG. 3. Box plots illustrating the response magnitude (the top row) and the latencies (the bottom row) as a function of phrase type. Box definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge−low hinge), and low hinge −1.5 (high hinge−low hinge) (Data Desk; Data Description). Asterisk bracket indicates significance between measures.

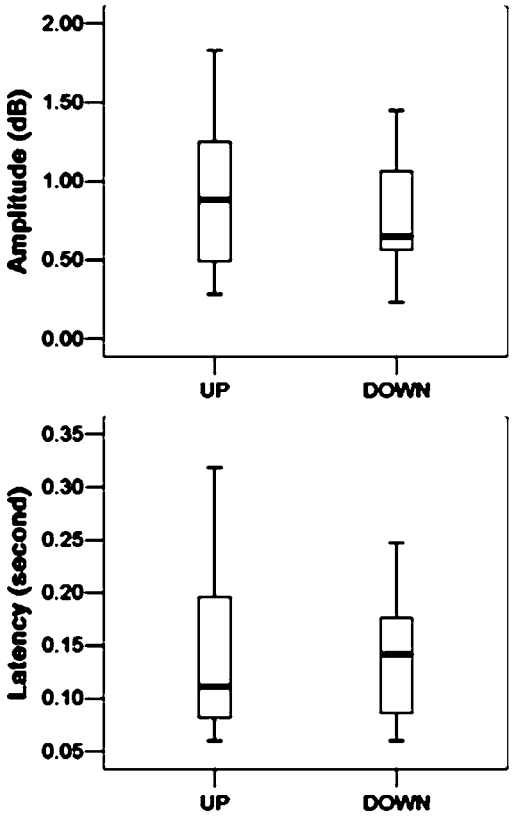

FIG. 4. Box plots illustrating the response magnitude (the top row) and the latencies (the bottom row) as a function of stimulus direction.

cus and No Focus were not significantly different. No significant interaction was found between phrase type and stimulus direction.

For response latency, two-way repeated-measures ANOVAs revealed significant main effects of phrase type ($F(2,18)=4.824$, $p=0.021$). Post hoc Bonferroni tests indicated that the latencies for the 1st-Focus phrase were significantly longer than those for the No-Focus phrase ($p=0.039$) (see Fig. 3 and Table V). Latencies for the 2nd-Focus and No-Focus conditions were not statistically different. Response latencies did not differ significantly as a function of stimulus direction.

IV. DISCUSSION

It is known that amplitude variation in speech is highly dependent on the characteristics of the speech sounds (Fant, 1960; Stevens, 1998) as well as lung volume (Ladefoged and Loeb, 2002). For example, other things being equal, amplitude is negatively related to the height of the vowel (hence positively related to F1)—the higher the vowel, the lower the amplitude. Also, amplitude is closely related to speech sound type: it is much higher in vowels than in consonants. These facts make it difficult for amplitude to be a direct object of control in speech production, as is evident in the findings that amplitude is not highly effective in encoding linguistic contrasts (Fry, 1958; Turk and Sawusch, 1996). Thus the question as to whether human subjects would exhibit compensatory responses to amplitude perturbation similar to those to pitch perturbation is highly interesting. The similarity in responses to loudness perturbation found in the present study and responses to pitch perturbations found in previous studies (Chen *et al.*, 2007; Xu *et al.*, 2004) suggest the existence of a highly automatic feedback mechanism that assists in the control of various aspects of the vocal output, including both F_0 and amplitude of voice.

The present study was designed to investigate vocal responses to loudness-shifted feedback during the production of Mandarin speech. In this paradigm, subjects varied their focus across three different phrases. At the same time, their voice loudness feedback was experimentally manipulated for

TABLE IV. Average response magnitudes (SD) across three phrase types and two stimulus directions.

Phrase		1st Focus	2nd Focus	No Focus
Direction	Up	1.07 (0.50)	0.66 (0.34)	0.99 (0.51)
	Down	1.02 (0.54)	0.67 (0.37)	0.77 (0.24)

TABLE V. Average response latencies (SD) across three phrase types and two stimulus directions.

Phrase		1st Focus	2nd Focus	No Focus
Direction	Up	171 (56)	168 (106)	100 (49)
	Down	159 (54)	130 (70)	135 (55)

TABLE VI. Average amplitude (SD) of the two syllables across three phrase types.

Phrase		1st Focus	2nd Focus	No Focus
Syllable	Syllable 1	84.62 (4.2)	83.85 (3.38)	82.27 (3.11)
	Syllable 2	81.16 (4.10)	85.13 (3.94)	82.26 (3.17)

short durations (200 ms). The 3 dB perturbations were added to or subtracted from the amplitude of the subjects' productions and were clearly distinguishable from the nonperturbed, control trials (see Fig. 1). The onset of the perturbation was presented usually during the first syllable, but for some subjects who increased their rate of speech, the perturbation onset occurred during the transition between the first and second syllables. In all cases the perturbation ended during the second syllable, and the response also occurred during the second syllable.

Results showed that, similar to pitch-perturbation studies, the subjects responded to increasing and decreasing loudness perturbations by changing their voice amplitude in response to the stimuli. Half the responses were compensatory, i.e., they were opposite in direction to the stimulus, and close to one third of the responses "followed" the direction of the stimulus. It is unknown why the number of "following" responses was greater than in previous studies of voice pitch or loudness feedback (Chen *et al.*, 2007; Xu *et al.*, 2004), particularly since it is not known what causes such responses in the first place. In a previous paper on responses to loudness perturbations during vowel phonations, greater numbers of "following" responses were speculated to result from the fact that the stimuli (−3 dB, 200 ms duration) were difficult to perceive (Larson *et al.*, 2007). This speculation is supported by the report that 100% of opposing responses were found in the case of 10 dB stimuli during vowel productions (Heinks-Maldonado and Houde, 2005). Also the fact that 33% of the responses were "following" in the present study while only 19% (Sivasankar *et al.*, 2005) and 20% (Larson *et al.*, 2007) were "following" for sustained vowels indicates that it may be more difficult to perceive the direction of short duration loudness perturbations during speech production compared to vowel phonations. Another possible explanation was that the subjects may have used the feedback signal as their choice of referent when making comparisons with their intended voice amplitude production. Using the feedback signal as the referent, as in matching a piano note while singing would cause a "following" response (Hain *et al.*, 2000). Each of these explanations may also apply to the results of the present study. The important, but still unanswered question regarding these responses is why subjects produce them under certain conditions and not others.

The magnitudes of the compensatory responses to perturbed loudness feedback during speech imply that the response magnitude is dependent on the relative voice amplitude at the moment of the stimulus. This is seen in the finding that responses for the 1st-Focus phrase were significantly larger than those for the 2nd-Focus phrase. As can be seen in Table VI, in the 1st-Focus phrases the mean amplitude of the first syllable was 3.46 dB higher than that of the second syllable, during which the response peaked in amplitude. In the 2nd-Focus phrases, the mean amplitude of the first syllable was 1.28 dB *lower* than that of the second syllable. Thus it is possible that the larger amplitude of the second syllable in the 2nd-Focus phrases may have partially inhibited the compensatory mechanism, as the underlying amplitude is rapidly rising during the response to the perturbation. It is also possible that, because the first syllable is shortened when focus is on the second syllable, as can be seen in Fig. 2, there is greater overlap of the stimuli with the second syllable than in the 1st-Focus condition. This would have increased the likelihood that the response peaked while the underlying amplitude was still increasing, thus affecting the accuracy of the response measurement. Regardless, the evidence is sufficient that in Mandarin, auditory feedback in the form of both pitch and loudness feedback is used both during the production of lexical tones (Xu *et al.*, 2004) and focus in speech.

Comparison of results between this and previous studies on pitch- and loudness-shifted voice feedback suggests similarities in the mechanisms underlying the responses. Specifically, response magnitudes to pitch- or loudness-shifted feedback were less than the stimulus magnitudes. Although direct comparisons between responses to pitch-shifted and loudness-shifted feedback are not possible because of their different acoustical dimensions, a rough comparison can be made if response magnitudes are treated as a percent of the stimulus (hereafter "% response magnitude"). In several pitch-shift studies, % response magnitudes for a 100-cent stimulus varied from 10% to 30% (Bauer and Larson, 2003; Burnett *et al.*, 1998; Hain *et al.*, 2000). With a pitch-shift stimulus of 25 cents, % response magnitudes approached 100% (Larson *et al.*, 2001). For the two previous studies of loudness-shifted feedback, % response magnitudes varied from 0.06% (0.61 dB for a ±10 dB stimulus) (Heinks-Maldonado and Houde, 2005) to about 90% (0.9 dB for a ±1 dB stimulus) (Bauer *et al.*, 2006). In the present study, the 2nd-Focus condition yielded the lowest mean response magnitude of 0.66 dB or 22% response magnitude, while the 1st-Focus condition yielded the highest, 1.07 dB SPL or 36% response magnitude. In both pitch- and loudness-shift studies, the largest % response magnitudes occurred with the smallest stimuli. As stimulus magnitude increased, % response magnitude decreased. This general finding suggests that responses to perturbed auditory are optimally suited to correct for small variations in voice pitch or loudness feedback.

Another similarity between this and previous studies is the fact that response magnitudes and latencies varied as a function of the vocal task. Natke *et al.* (2003) demonstrated larger responses to pitch-shifted feedback during singing compared to speech. Xu *et al.* (2004) demonstrated larger responses and shorter latencies to pitch-shifted feedback when the stimulus was presented prior to a major change in the tone, e.g., the transition from a high to a falling tone. Chen *et al.* (2007) showed that in English speech, larger and quicker responses to pitch-shifted feedback occurred when downward pitch perturbations were presented prior to a rise

in voice F_0. In the present study, responses to the loudness-shifted feedback were significantly larger for the 1st-Focus phrase than those for the 2nd-Focus phrase. In addition, latencies were significantly longer in the 1st-Focus than in the No-Focus pattern. Thus, the modification of response magnitudes and latencies in both pitch- and loudness-shifted feedback studies during speech, indicates the nervous system is capable of modulating the influence of auditory feedback for the control of the voice.

However, even though the present study demonstrated task-dependent modulation of voice amplitude responses to loudness-shifted feedback, the nature of these modulations is quite different than those reported in earlier pitch-shift studies. In previous pitch-shift studies, latencies generally became shorter as response magnitude increased (Chen *et al.*, 2007; Xu *et al.*, 2004). In contrast, in the present study, the larger responses in the 1st-Focus condition also had longer latencies. Similarly, the shorter latency responses in the No-Focus condition did not have larger magnitudes. Thus, with pitch-shifted feedback during speech, as responses increased in magnitude, they also became quicker. With loudness-shifted feedback, however, as responses became larger, they became slower. These differences between simultaneous changes in response magnitude and latency may suggest fundamental differences in the way the nervous system uses voice loudness feedback for control of voice amplitude vs. pitch feedback for control of voice F_0 during speech. Further research should be addressed to this difference.

V. CONCLUSION

Results of the present study showed that native speakers of Mandarin made compensatory responses to loudness perturbations in a manner similar to that of previously reported responses to pitch-shifted feedback, with a time delay of about 142 ms, and a magnitude about 29% of the stimulus magnitude. The finding is highly significant given that the role of amplitude variation in speech is known to be not nearly as effective as that of pitch variation (Fry, 1958; Turk and Sawusch, 1996). Compensatory responses to perturbed auditory feedback thus seem to be part of a mechanism that reacts quite automatically to any discrepancy between the anticipated and actual feedback. Results also showed that the response magnitude was smaller if the perturbation occurred when the linguistic focus was utterance final than when it was nonfinal. This provides further evidence that auditory-feedback control of vocalization is dependent on the nature of the speech signal at the time of the perturbation (Chen *et al.*, 2007; Xu *et al.*, 2004). Future studies will help to define which specific speech segments are more or less sensitive to auditory feedback.

ACKNOWLEDGMENTS

This study was supported by a grant from NIH, Grant No. DC006243-01A1. We thank Chun Liang Chan for his help with computer programming. Portions of this manuscript were presented at a meeting of the Acoustical Society of America in May, 2006, Providence, RI.

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.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

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.

Chang-Yit, R., Pick, H. L., and Siegel, G. M. (**1975**). "Reliability of sidetone amplification effect in vocal intensity," J. Commun. Dis. **8**, 317–324.

Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F0 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.

Draper, M. H., Ladefoged, P., and Whitteridge, D. (**1959**). "Respiratory muscles in speech," J. Speech Hear. Res. **2**, 16–27.

Dreher, J. J., and O'Neill, J. J. (**1958**). "Effects of ambient noise on speaker intelligibility of words and phrases," Laryngoscope **68**, 539–548.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.

Garber, S. F., Siegel, G. M., and Pick, H. L. (**1976**). "The influence of selected masking noises on Lombard and sidetone amplification effects," J. Speech Hear. Res. **19**, 523–535.

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.

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

Hirano, M., and Ohala, J. (**1969**). "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles," J. Speech Hear. Res. **12**, 362–373.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Isshiki, N. (**1964**). "Regulatory mechanism of voice intensity variation," J. Speech Hear. Res. **7**, 17–29.

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.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). "Loudness predicts prominence: Fundamental frequency lends little," J. Acoust. Soc. Am. **118**, 1038–1054.

Koyama, T., Kawasaki, M., and Ogura, J. H. (**1969**). "Mechanics of voice production. I. Regulation of vocal intensity," Laryngoscope **LXXIX**, 337–354.

Ladefoged, P., and Loeb, G. (**2002**). "Preliminary studies on respiratory activity in speech," UCLA Working Papers in Phonetics **101**, 50–60.

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

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., and Perkell, J. (**1995**). "Phonemic resetting versus postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time," J. Acoust. Soc. Am. **98**, 3096–3106.

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., 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 F0 and amplitude control," J. Acoust. Soc. Am. **121**, 2862–2872.

Leder, S. B., Spitzer, J. B., Milner, P., Flevaris-Phillips, C., Kirchner, J. C.,

and Richardson, F. (**1987**). "Voice intensity of prospective cochlear implant candidates and normal hearing adult males," Laryngoscope **97**, 224–227.
Liu, F., and Xu, Y. (**2005**). "Parallel encoding of focus and interrogative meaning in mandarin intonation," Phonetica **62**, 70–87.
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.
Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.
Siegel, G., and Pick Jr., H. L. (**1974**). "Auditory feedback in the regulation of voice," J. Acoust. Soc. Am. **56**, 1618–1624.
Siegel, G. M., and Kennard, K. L. (**1984**). "Lombard and sidetone amplification effects in normal and misarticulating children," J. Speech Hear. Res. **27**, 56–62.
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.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
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**, 1284–300.
Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.
Turk, A. E., and Sawusch, J. R. (**1996**). "The processing of duration and intensity cues to prominence," J. Acoust. Soc. Am. **99**, 3782–3790.
Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
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.