Journal of Phonetics (2002) **30,** 303–320
doi:10.1006/jpho.2001.0160
Available online at http://www.idealibrary.com on IDEAL®

The role of auditory feedback during phonation: studies of Mandarin tone production

Jeffery A. Jones*

ATR-International Human Information Science Laboratories, Communication Dynamics Project, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan and Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6

K.G. Munhall

Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6 and Department of Otolaryngology, Queen's University, Kingston, Ontario, Canada K7L 3N6

Received 9th November 2001, and accepted 16th December 2001

Auditory feedback during speech plays an important role in the control of articulation. The sound of a speaker's voice is used to calibrate speaking volume and vocal pitch and also influences the precision of articulation. In two experiments, the speech feedback system is studied by exposing subjects to modified auditory feedback regarding their vocal pitch. In previous work, Jones & Munhall (2000) demonstrated that short-term exposure to altered feedback regarding vocal pitch led to aftereffects in pitch production when feedback was returned to normal. The adaptation suggests that a remapping between produced pitch and expected feedback occurred. In the first study, native speakers of Mandarin were exposed to sudden pitch feedback perturbations while producing /ma/ as a high, flat tone. All subjects showed rapid compensation in response to the perturbation. In the second study, Mandarin speakers were asked to say the same tonal stimulus in two experimental conditions. In one condition, auditory feedback regarding their F_0 was slowly shifted up one semitone without their awareness. In a contrasting condition, their feedback was slowly shifted down one semitone. Results indicate that subjects compensated for the pitch-shifted feedback and negative aftereffects were observed when feedback was suddenly returned to normal after the short-term exposure to the altered feedback. These results parallel those found for English speakers and suggest that despite having an obligatory underlying phonetic representation, the acoustic-motor representation is malleable.

*Address correspondence to J. A. Jones, ATR-International Human Information Science Laboratories, Communication Dynamics Project, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan. E-mail: jones@atr.co.jp

0095–4470/(02)/$ - see front matter

1. Introduction

One of the most enduring questions in the study of speech production is the role played by sensory feedback in motor planning and control. In this paper, we focus specifically on auditory feedback and the variety of different roles it plays in speech motor control. Access to auditory feedback from your own speech is particularly important for developing normal speech production. Partial or total deafness that occurs early in life can severely affect the quality of a child's articulation or even prevent the normal development of speech altogether (Smith, 1962; Oller & Eilers, 1988). If hearing loss occurs after language acquisition, the deleterious effects on speech are greatly reduced in comparison (Waldstein, 1990; Cowie & Douglas-Cowie, 1992).

Several lines of evidence demonstrate that auditory feedback still remains necessary for accurate speech production in adults. Difficulties in controlling the pitch and intensity of speech are relatively common in adults who are deafened and occur rather soon after deafness onset (Binnie, Daniloff & Buckingham, 1982; Waldstein, 1990; Lane & Webster, 1991; Cowie & Douglas-Cowie, 1992). In addition, a reduction in intelligibility often occurs because imprecise production of consnants and vowels occurs after long periods without the benefit of audition (Waldstein, 1990). Normal hearing speakers in noisy conditions attempt to regain the advantages of auditory feedback by compensating for the noise and increasing the loudness and the duration of their utterances (Ringel & Steer, 1963; Lane & Tranel, 1971).

Various laboratory studies have also shown that altering auditory feedback affects speech production. When auditory feedback is delayed, dramatic disruptions in speech production are common (Smith, 1962). Selectively filtering frequencies of auditory feedback can cause subjects to change their productions depending on the nature of the filtered spectra (Garber & Moller, 1979). In addition, a number of studies have demonstrated that when feedback regarding F_0 is suddenly raised or lowered, subjects compensate by shifting their voice fundamental frequency in the opposite direction (e.g., Kawahara, 1995; Burnett, Freedland, Larson & Hain, 1998; Jones & Munhall, 2000).

These developmental and adult data suggest that auditory feedback plays a key role in the creation and maintenance of some form of internal speech motor representation. During development, auditory feedback provides information necessary to represent speech motor goals and it aids error correction during the learning of these representations. In mature speech, auditory feedback is important for maintaining the precision of articulation. It does so by its involvement in closed-loop control of the articulators but also through the calibration of the internal representations used in articulatory control.

This conclusion is consistent with a growing body of work in general motor control that suggests that the nervous system relies on "internal models" of movement to support rapid skilled movement. Internal models are conceptualized as neural representations of the kinematic and force parameters of a movement as well as the expected proprioceptive information. The models may be used by the nervous system to predict movement outcome and provide internal feedback to planning and control systems. Having an internal model that provides

information to an internal feedback loop effectively avoids the delays associated with reliance on peripheral feedback (Wolpert, Ghahramani & Jordan, 1995; Desmurget & Grafton, 2000).

A number of recent studies have supported the existence of internal models. For example, Flanagan & Wing (1993) have shown that the force with which an object is gripped changes in synchrony with changes in load forces on the object. The control system appears to predict the loads on the object and adjusts grip force based on the inertial forces expected. Similar findings have been found when grip adjustments must be made based on the surface textures of the object (Johansson & Westling, 1984).

Evidence that internal models play a role in internal feedback loops comes from studies of oculomotor control. When the eye is perturbed as the oculomotor system prepares or is in the middle of making a saccade to a flashed target, a compensatory saccade reestablishes gaze back to the target location stored in memory (Pelisson, Guitton & Goffart, 1995; Keller, Gandhi & Shieh, 1996). The compensatory saccades occur even in the absence of visual or proprioceptive feedback in monkeys (Guthrie, Porter & Sparks, 1983). A feedforward model providing internal feedback to the oculomotor control system can account for these observations.

There is recent evidence suggesting that speech production may utilize internal models for control. Houde & Jordan (1998) asked subjects to whisper one-syllable words while receiving auditory feedback in which the formants they were producing were shifted enough to change the vowel's phonetic identity. After a short period of exposure to the altered feedback, subjects compensated for the formant transformations. The modified productions persisted even in the absence of feedback. The results imply that the mapping between the vocal tract movements and the consequent acoustics were modified.

The adaptability of the articulatory system has been demonstrated using a variety of different perturbation paradigms. When unexpected mechanical loads are applied to the lower lip, rapid compensatory changes in the lips and jaw (Abbs & Gracco, 1984; Gracco & Abbs, 1988) and the larynx (Löfqvist & Gracco, 1991; Munhall, Löfqvist & Kelso, 1994) have been observed. When the jaw is loaded, compensatory changes in the lips (Folkins & Abbs, 1975; Shaiman, 1989), tongue (Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984) and velum (Kollia, Gracco & Harris, 1992) have been measured. More static, structural perturbations to the vocal tract have also been tested. In a series of studies, Hamlet and colleagues (e.g., Hamlet & Stone, 1976, 1978; Hamlet, Stone & McCarty, 1978) had subjects learn to speak while wearing a dental prosthesis that thickened the palate in the alveolar ridge region. Recently, McFarland, Baum & Chabot (1996) have also shown that subjects can adapt to a new artificial vocal tract morphology and Sorokin, Olshansky & Kozhanov (1998) have investigated adaptation to surgically modified vocal tracts. In an innovative combination of sudden mechanical perturbations and static structural manipulations, Honda & Fujino (2000) have shown that subjects can overcome an unexpected change in the palatal shape. Interestingly, Honda & Fujino found that the presence or absence of auditory feedback influences the accuracy of articulation adaptation to a palatal change.

Recently, we have addressed the role of auditory feedback by testing whether altered feedback regarding pitch could change an acoustic-motor representation used

in F_0 control (Jones & Munhall, 2000). Previous studies have shown that subjects compensate when auditory feedback regarding their own pitch is suddenly raised or lowered artificially (e.g., Elman, 1981; Kawahara, 1995; Burnett *et al.*, 1998). These compensatory responses have been interpreted as support for the idea that F_0 control is reliant on feedback. In our experiment, vocal pitch feedback was slowly shifted up or down in frequency while English-speaking subjects produced vowels. Despite being unaware of the feedback manipulation, subjects modified their produced F_0 in the opposite direction of the shifted feedback, confirming the observations of other pitch perturbation studies. When F_0 was shifted up, subjects lowered their pitch and when it was shifted down they raised their pitch. In addition, the subjects also showed evidence of sensorimotor adaptation. When F_0 feedback was returned to normal after exposure to the altered feedback conditions, significant aftereffects were observed. After a short period of hearing their F_0 feedback shifted higher than normal, the pitch of their voice increased when they were unexpectedly given normal feedback. Conversely, subjects decreased their pitch when they were given normal feedback after hearing their voice shifted down in frequency. The adaptation indicates that an internal model or representation may play a role in the long-term calibration of vocal pitch. That is, auditory feedback may be used to update an internal representation of the mapping between pitch output and the motor systems that control it.

Perkell and colleagues (Svirsky, Lane, Perkell & Webster, 1992; Perkell, Matthies, Lane, Guenther, Wilhems-Tricaria, Wozniak & Guiod, 1997) have suggested that the role played by auditory feedback depends on whether the feedback is related to postural or phonemic control. Postural variables in their terminology include speaking rate, sound level and F_0 and these parameters change rapidly in response to changes in feedback. Phonemic parameters include vowel formant frequency, voice onset time and the spectral characteristics of consonants. These parameters are more stable in the face of feedback changes. Partial support for this distinction comes from the differential stability of acoustic speech parameters following the onset of deafness (Cowie & Douglas-Cowie, 1992). Speaking volume, rate and pitch are more severely and rapidly impaired by postlingual deafness than are the acoustics of individual consonants and vowels. The studies of cochlear implant patients by Perkell *et al.* also indicate that these postural parameters are very sensitive to changes in hearing status.

A number of possible explanations can be offered for this phonemic/postural contrast. As suggested by Perkell *et al.* (1997), this contrast may reflect different reliance on feedback for different aspects of speech. Speaking level and vocal pitch may require more continuous tuning to speaking conditions, whereas the accuracy of consonant and vowel spectra may be more optimally served by a strong internal representation. Another possibility is that multiple internal models with different temporal characteristics may contribute to control of the articulators in speech (see Wolpert & Kawato, 1998). According to this explanation, all articulation involves internal models but different aspects of speech planning and control are separately represented. Such an idea is consistent with the segmental/suprasegmental distinction and the distributed neural representation for these different aspects of speech (Baum & Pell, 1999). Finally, it is possible that individual articulators are controlled in different manners. Vocal pitch and loudness have significant laryngeal involvement

and the larynx may have a unique control system compared to supraglottic articulators.

One way of examining some of these hypotheses is to hold the articulator (the larynx) constant and manipulate the role played by vocal pitch feedback. This can be done by exposing subjects who have learned to produce linguistically contrastive pitch targets to the altered feedback paradigm. In English, vocal pitch varies during normal conversation depending on prosodic pattern, emotionality, as well as the intensity and rate of speech (Zemlin, 1981). While an individual may have a preferred or habitual vocal pitch around which their conversational pitch varies (Coleman & Markham, 1991), the pitch contour within a vowel is not used contrastively. On the other hand, tone languages use vocal pitch targets to distinguish words and grammatical categories.

The phonological status of tone has changed considerably in the last 25 years (Goldsmith, 1976; Clements, 1985; Clements & Hume, 1995; Odden, 1995; Yip, 1995) with a consensus supporting an ''autosegmental'' view of tone representation. In this approach, tone and some other features (see Clements & Hume, 1995) are viewed as separate levels of representation (autosegments) that are independent of the segments in an utterance. Taking this approach to the study of tone languages solved a number of problems such as how to represent dynamic tone contours and how to account for tone preservation following vowel deletion. From our perspective, however, the intricacies of the phonological representation of tone are less important than the perceptual role and perhaps the production precision of fundamental frequency for tone targets. Although, as often is the case in English, context helps a listener identify the intended meaning of words, vocal pitch is crucial for unambiguous perception of meaning in tone languages.

Altering the auditory feedback for speakers of a tone language thus tests the role of auditory feedback for the laryngeal control of a setting that is not simply postural. In a tone language there is an explicit pitch target for the talker to ''aim'' for. While tones vary in absolute pitch because of intonation contour, stress, and sandhi (Kratochvil, 1998; Wu, 2000), the primary perceptual cue used to distinguish different tone categories is the fundamental frequency contour within a vowel (Gandour, 1978). Thus, fundamental frequency in tone languages plays a different role than in a language such as English. Vocal pitch in English corresponds directly to Perkell *et al.*'s postural settings while in tone languages, vocal pitch plays this role as well as a more phonemic role.

In this paper, we report data from native speakers of Mandarin. Mandarin is a tone language in which there are four tones (and a neutral tone). The four tones are (1) mid-high flat, (2) rising from mid-low to mid-high, (3) Low flat with rising end and (4) falling from high to low (Gandour, 1978). Each morpheme in Mandarin is monosyllabic and the association of different tone contours with a morpheme distinguishes a different meaning. For example, ''ma'' when pronounced as tones 1–4 means ''mother'', ''hemp'', ''horse'' and ''to scold'', respectively (Gandour, 1978).

In Experiment 1, we test whether speakers of Mandarin compensate when unexpected, large pitch perturbations are introduced to their auditory feedback during the production of an utterance. In Experiment 2, we replicate Jones & Munhall (2000) with Mandarin speakers. In this second experiment, we test for compensation to small incremental changes in fundamental frequency and then test for evidence of adaptation to this transformed feedback.

2. Experiment 1

A number of studies have demonstrated that subjects who receive sudden pitch feedback perturbations quickly modify their F_0 in response (Elman, 1981; Kawahara, 1995; Larson, 1998). These studies have included subjects speaking nontone languages (English and Japanese). In this first experiment, we wanted to determine if subjects producing Mandarin tones similarly modified their F_0 productions when exposed to altered pitch feedback. If speakers of tone languages do not rely as heavily on auditory feedback to control their pitch, we would not expect to see rapid pitch modifications when feedback is perturbed.

2.1. *Method*

2.1.1. Subjects

Twelve speakers (six women and six men) whose first language was Mandarin participated. The subjects were between 18 and 35 years of age (mean of 27.5 years), grew up in Mandarin-speaking communities and received their primary education in Mandarin. At the time of the experiment, they were all in Canada studying at Queen's University. The subjects reported no hearing, speech, or language problems.

2.1.2. Apparatus and procedure

The experimental setup is shown in Fig. 1. The recording sessions took place in a double-walled soundproof booth (Industrial Acoustics Corporation, Model 1204).

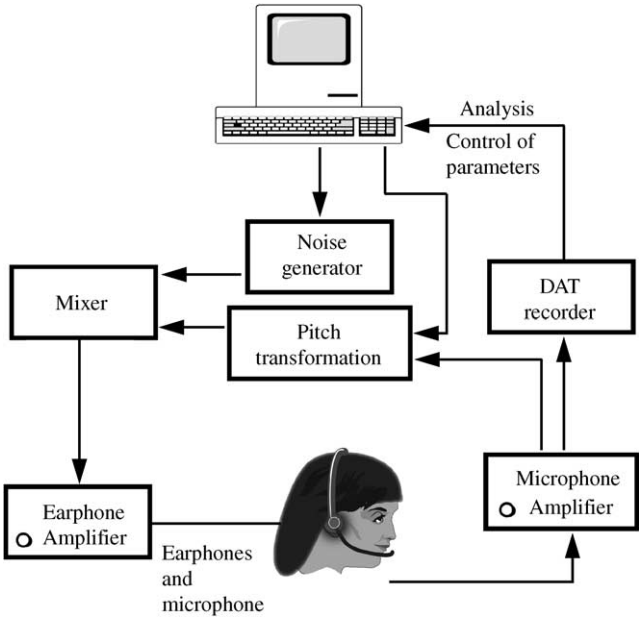

Figure 1. Schematic of the acoustic feedback setup.

The subjects were seated in front of a computer monitor. On the computer screen, the Mandarin word for mother was depicted in traditional Chinese script (pronounced "ma" with the Mandarin high, flat tone). Underneath the ideogram was a countdown from 2 to 0 s. Subjects pronounced the word on the screen for the duration of the countdown and then pressed a mouse button to initiate the next trial. The subjects were not informed that their feedback would be manipulated in any way. Subjects were merely instructed to produce the word as similarly as possible on each trial.

While saying the words, subjects wore a headset microphone (Shure WH20) that was maintained at a fixed distance from the mouth ($\sim$7 cm). The microphone signals were amplified (Tucker-Davis MA2 microphone amplifier) and filtered (Tucker-Davis FT6-2) with a 9 kHz frequency cutoff. The speech signals were then fed into an Eventide Ultra Harmonizer (H3000-D/SX)[1] which shifted the pitch of the signals. The processing necessary for the pitch-shift introduced a very small delay of 3 to 4 ms. The frequency-altered speech was then mixed with pink noise (Grason-Stadler 901B) and a multi-speaker babble (Auditec, St. Louis, MO). Both the signal and noise were amplified by a Yorkville reference amplifier (model SR 300) which transmitted the sound through Etymotic (ER-2) earphone's foam inserts positioned comfortably in the subject's ear canals. The level of the masking noise (pink and multi-speaker babble) was approximately 75 dB SPL and was used to decrease the amount of natural acoustic feedback that the subjects received regarding their true vocal pitch.

Subjects produced 100 utterances. Ten utterances were produced at the beginning and the end of the session and were not altered. The first 10 trials were used to acclimate the subjects to the task. After these trials, subjects produced the word on 40 control trials and 40 experimental trials. During the experimental trials, auditory feedback regarding the subjects' pitch was suddenly shifted up one semitone. On control trials the feedback was not altered. The control and experimental trials occurred pseudo-randomly so that subjects could not predict on which trials their feedback might be altered. In addition, the perturbations occurred randomly during an utterance (between 1000 and 1500 ms after trial initiation). Trial initiation and the pitch processor were controlled by a computer. Each session was recorded with a sampling frequency of 48 kHz on DAT. The data were then low-pass filtered (with a 5 kHz cutoff) and digitized with a sampling rate of 11.025 kHz. F_0 contours for the utterances during each trial were determined using an algorithm included in the Praat software program (Boersma, 1993).

2.2. *Results*

F_0 contours for each of the utterances produced on the 40 perturbation and 40 control trials were calculated. Fig. 2 shows an F_0 contour in Hertz of a typical subject for a single perturbation trial. Out of 400 perturbation trials, 40 trials were eliminated from the analysis because subjects did not begin speaking

[1]The pitch-shifting algorithm used in the Harmonizer is not publicly distributed but uses a waveform-based method to change fundamental frequency. This hardware has been used in a number of previous experiments (e.g., Burnett, Senner & Larson, 1997; Burnett *et al.*, 1998; Jones & Munhall, 2000; Kawahara, 1995).

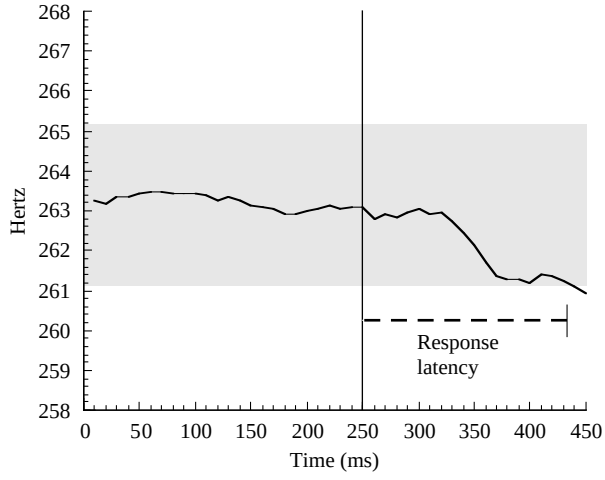

Figure 2. An F_0 contour for a subject's perturbation trial. The vertical line at 250 ms is the time onset of the pitch-shift stimulus. Any pitch that lies within the shaded area is ± 2 S.Ds from the prestimulus mean F_0. A response is counted when an F_0 deviates ± 2 S.Ds.

TABLE 1. The mean, standard error and median response latency (ms) for the 10 subjects in Experiment 1

Subject	Mean	S.E.	Median
1	230.36	37.43	150
2	194.67	30.63	135
3	237.50	36.86	155
4	185.95	19.33	130
5	194.06	24.37	130
6	258.09	39.81	160
7	182.65	18.16	160
8	202.38	37.87	140
9	237.42	39.78	150
10	186.00	34.18	100
Average	210.91	31.84	141

more than 250 ms before a perturbation and 15 because subjects did not show a response before 1000 ms elapsed. In addition, 18 trials were not included due to F_0 tracking errors. For each perturbation trial, the standard deviation of F_0 was calculated for the 250 ms prior to the perturbation onset. A response to the altered feedback was counted if the subject's pitch deviated at least ± 2 S.D.s from the prestimulus mean. All subjects made compensatory responses, that is on average they lowered their pitch in response to feedback perturbation. The response latency varied from trial to trial but the mean response time was quite rapid (see Table 1). Mean response latency across subjects was 211 ms after the onset of feedback alteration. The average response latency for each subject varied between 183 and 258 ms.

2.3. *Discussion*

The results in Experiment 1 demonstrate that speakers of Mandarin compensate when feedback regarding the pitch of their tone is suddenly increased. All subjects showed rapid compensation in vocal pitch by lowering their F_0 when pitch feedback was raised. The study replicates the findings of a number of studies that show speakers of English and Japanese compensate when exposed to altered feedback paradigms (Elman, 1981; Larson, Burnett, Kiran & Hain, 2000; Kawahara, 1995). Larson *et al.* (2000) maintain that the audio-vocal system controls F_0 production using auditory feedback in a closed-loop fashion. The data in Experiment 1 provide further support for their closed-loop negative feedback model using Mandarin subjects.

It should be noted that all the subjects in this study compensated for the pitch-shifted feedback they heard. Larson and his colleagues have had difficulties accounting for a small but reliable number of English subjects that consistently "follow" the pitch-shifted feedback as opposed to compensating for it (Burnett *et al.*, 1997). The fact that our subjects all compensated and did not follow during exposure to the altered feedback could be due to methodological differences between laboratories or simply a chance occurrence specific to our particular sample. However, it may instead indicate a difference regarding the use or ability to use feedback in Mandarin speakers. Xu & Sun (2000) have recently reported that Mandarin speakers differed from English speakers in the speed at which they changed their F_0 productions from one pitch target to another. Subjects were asked to produce a quick succession of high and low pitches by imitating the pitch changes they heard in a synthesized stimulus. English speakers were found to produce larger undulations which were correlated with faster pitch changes. The differences in performance in their study may indicate a language-difference in the ability to use the auditory stimulus target or the ability to produce the necessary productions accurately. However, the mean and variance of the response latencies observed in our data are similar to those reported for English speakers using a similar methodology (e.g., Burnett *et al.*, 1998; Larson *et al.*, 2000). This similarity is not consistent with a language-based difference.

3. Experiment 2

We have recently shown that after relatively short periods of hearing their auditory feedback shifted either up or down in pitch, English-speaking subjects show aftereffects when their feedback is returned to normal (Jones & Munhall, 2000). These aftereffects suggest that exposure to the altered feedback conditions resulted in modifications of an internal model or representation of the mapping between pitch output and the motor systems that control it. The control of vocal pitch in English is dependent on what Perkell *et al.* (1997) refer to as postural settings. The pitch contour within a vowel is not used contrastively to distinguish between words or grammatical categories. Rather, pitch plays a much more phonemic role in Mandarin. In Experiment 2, we replicate our previous work with Mandarin-speaking subjects. If the internal representation of pitch targets is more robust in Mandarin speakers, then speakers will be less likely to show aftereffects even if they do compensate as shown by Experiment 1.

3.1. *Method*

3.1.1. Subjects

Ten female speakers whose first language was Mandarin participated in this study. They were between 18 and 25 years of age (mean of 21.5 years). None of the subjects had previously served in a speech feedback experiment. Like the participants in Experiment 1, the participants grew up in Mandarin-speaking communities until moving to Canada to study. The participants reported that they did not have hearing, speech, or language problems.

3.1.2. Apparatus and procedure

The apparatus and setup were the same as that used in Experiment 1 (see Fig. 1). There were two experimental conditions in the study. In a "Down" condition, the pitch of the subject's auditory feedback was slowly decreased during the session. On a subsequent day, the same subjects participated in an "Up" condition where the feedback was slowly increased. The sessions occurred on separate days in order to reduce vocal fatigue.

On the computer screen in front of each subject was the Mandarin word for mother in traditional Chinese script (pronounced "ma" with the Mandarin high, flat tone). Subjects pronounced the word on the screen for 2 s. Subjects were asked to produce the word as similarly as possible on each trial. Each session consisted of three distinct phases. At the beginning of each session, subjects produced 10 utterances while receiving normal auditory feedback through the earphones. The mean F_0 value for these 10 initial utterances was taken as the subject's baseline F_0 for the session. This baseline phase was followed immediately by the training phase in which 120 utterances were produced. The pitch of the subjects' auditory feedback was either increased (Up condition) or decreased (Down condition) by 1 cent[2] for each successive utterance until the feedback received was 100 cents above the subject's true vocal pitch for the Up condition and 100 cents below for the Down condition. Immediately following these 100 trials were 20 trials in which feedback was held at 100 cents above or below the subjects' true F_0. After the training phase subjects produced 10 more utterances while receiving normal auditory feedback. The mean F_0 value for these 10 final utterances was taken as the subjects' final baseline F_0 and used to determine if aftereffects resulted from the training condition.

The Up and Down conditions were essentially the same apart from the direction of the pitch transformations. During the session, both the trial initiation and the pitch processor were controlled by a computer so that the phases of the experiment occurred without interruption or notice on the part of the subject.

3.2. *Results*

The pitch-shifts were intentionally very small so that subjects would have difficulty detecting them. A shift of 1 cent for a subject with an F_0 of 200 Hz is only 0.12 Hz. Indeed, in postexperimental interviews, subjects were surprised when told that the pitch of their voice had been altered. The initial F_0's under conditions of normal

[2]1 cent is equal to 1/100 of a semitone (see Apel, 1969).

feedback during the two experimental sessions (produced on different days) were essentially the same and only differed by 1.78 Hz (LSD test; $p = 0.68$). However, vocal pitches for the two conditions differed significantly during the last 20 trials of training, while subjects received pitch-shifted feedback (12.62 Hz; LSD test; $p = 0.008$). When subjects heard their F_0 100 cents higher than it actually was, they lowered their pitch in comparison to when they heard their F_0 100 cents lower than their true vocal pitch. When the subjects heard their feedback suddenly returned to normal, they raised their pitch in the Up condition by 13.3 Hz, and lowered their pitch in the Down condition by 8.1 Hz generating a significant interaction ($F\ (1, 9) = 161.6$, $p = 0.0000005$). However, the pitches during the final baseline trials were not quite significantly different (LSD test; $p = 0.052$).

In addition to analysis of the raw data in Hertz, the data were also normalized by converting them to cents. The experimental sessions took place on different days so in addition to the natural between-subject variation in F_0's there may have also been significant within-subject variation. That is, a subject may have had different baseline F_0's across the two sessions. The mean frequency value for each utterance during the training and final baseline were converted to cents and therefore normalized with respect to the F_0 baseline trials produced at the beginning of each session. These values were converted to cents using the formula:

$$\text{Cents} = 100\ (12\ \log_2 F/B)$$

In the formula, F is the mean frequency for the utterance during the trial. B is the grand mean frequency for the 10 utterances produced during the baseline phase at the start of the experimental session.

In Fig. 3(a), the mean F_0 for the final 20 trials of the training phase and the mean F_0 for the final baseline phase for both the Up and Down conditions is presented in cents. The same data in Hertz is depicted in Fig. 3(b) for direct comparison. The same significant interaction occurs before and after the data are normalized ($F\ (1, 9) = 146.3$, $p = 0.0000007$). When subjects were exposed to increased pitch feedback, they compensated and lower their vocal pitch and in comparison, did the opposite when they were exposed to decreased pitch feedback. When the feedback was returned to normal, subjects increased their pitch by 87.13 cents in the Up condition and decreased their pitch by 51.31 cents in the Down condition. During the final baseline trials, the two conditions differed significantly by 70.9 cents (LSD test, $p = 0.0000106$).

It can also be seen in Fig. 3(b) that in both experimental conditions, subjects produced F_0's above their initial baseline values during the training and final baseline phases (in both conditions, their pitch was above the baseline mean, that is above 0 cents). The finding that subjects increase their F_0 over the many trials is consistent with observations made by Jones & Munhall (2000). To ensure that the pitch changes observed in this study were not caused by changes in speaking volume over the course of a session, the root-mean-square (rms) amplitude of the utterances was calculated for the baseline trials, the final 20 training trials and the final baseline trials for the two conditions. There were no significant differences between the Up and Down conditions in the loudness of utterances ($F\ (1, 9) = 0.57$, $p = 0.47$). In addition, there was no significant main effect of phase in the sessions (baseline, training, final baseline: $F\ (2, 18) = 0.84$, $p = 0.45$) and no interaction between feedback

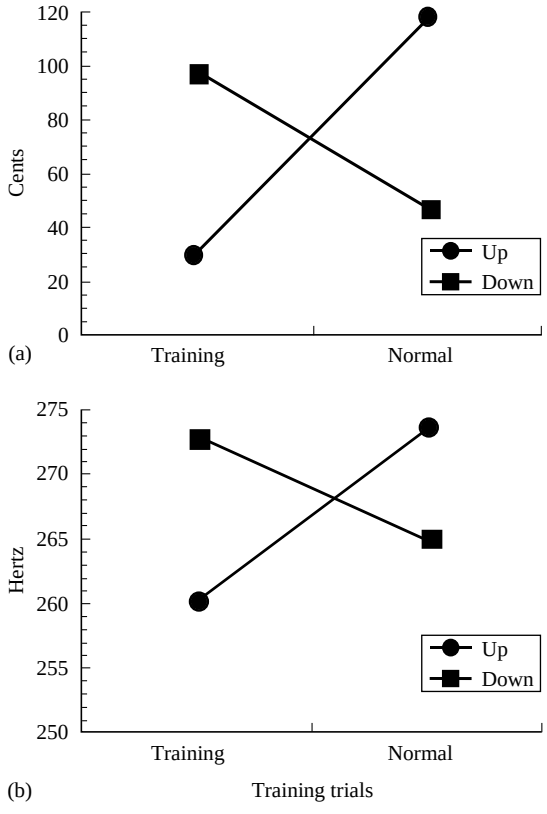

Figure 3. Average F_0 in cents (a) and Hertz (b) for the Up (circle) and Down (square) conditions during the final 20 trials of the training phase and when the feedback was returned to normal in the final baseline phase.

condition and phase (F (2,18) = 0.37, p = 0.69). The null effects indicate that speaking volume did not contribute to the observed changes in produced pitch in the study.

3.3. *Discussion*

Auditory feedback regarding the F_0 productions of Mandarin speakers was shifted up and down in pitch in this study. The results show two primary effects that resulted from exposure to the altered feedback. First, subjects compensated for pitch shifts by increasing their F_0 when they heard the pitch of their voice shifted down and by decreasing F_0 when it was shifted up. In addition, after the relatively short period of exposure to the frequency-altered feedback, subjects showed evidence of sensorimotor adaptation. When subjects heard their F_0 feedback shifted up for a period of time, they increased the pitch of their voice when they unexpectedly received normal, unaltered feedback. When they received feedback in which their F_0 was shifted down, the opposite effect was observed and they reduced the pitch of their voice when given normal feedback.

The pattern of adaptation shown here is similar to that found for English speakers. Although we did not include a condition without pitch-shifts, the interaction observed between the two pitch manipulations support the conclusion that some type of internal model or representation plays a role in the long-term calibration of vocal pitch—regardless of the role pitch plays in a particular language context.

The finding that subjects compensate when they hear alterations in their F_0 feedback is consistent with the results found in Experiment 1 as well as previous work. Most subjects have been found to compensate for sudden pitch-shifts in feedback by modifying their F_0 in the opposite direction (Burnett *et al.*, 1998; Kawahara, 1995). In the present study, very gradual and small shifts of 1 cent were involved (until the feedback was returned to normal). Consistent with our previous study of English subjects, these Mandarin subjects did not notice these individual perturbations. The Mandarin subjects compensated for these shifts in the same way that English speakers did. Taken together, the evidence from Experiments 1 and 2 indicate that auditory feedback can also be used in a closed-loop fashion to control fundamental frequency in speakers of tone and nontone languages.

Although significantly lower in pitch than the Down condition, the average pitch in the Up condition during maximum pitch-shift was actually higher than that observed during baseline phase at the start of the session. There was a tendency for subjects to gradually increase their pitch during the experimental session independent of the feedback condition. This tendency was also observed in our earlier study when English speakers were exposed to the same paradigm (Jones & Munhall, 2000). One possible explanation for this pattern is that because we did not ask subjects to maintain a particular loudness level, they increased their speaking volume during the session causing an associated rise in vocal pitch (see Gramming, Sundberg, Ternström, Leanderson & Perkins, 1988). However in both studies, there was no significant difference between the rms amplitude of the utterances produced at the beginning and end of the sessions ruling that factor out as a likely cause. Another possibility is that the increase in pitch was the result of vocal fatigue that developed over the session. We cannot test this hypothesis because there is no established method for assessing fatigue from an acoustic record (Titze, 1994). However, because the conditions were exactly the same apart from the direction of the pitch shift, the differential effects of the manipulations are evident.

It is worth noting again that as in Experiment 1, all subjects made compensatory responses and did not follow the pitch of the auditory stimulus. This finding is extraordinary given that a small percentage of English-speaking subjects are routinely observed to increase their vocal pitch in response to sudden, increased pitch-shifted feedback and decrease their pitch when exposed to sudden decreases in the pitch of auditory feedback. However, this difference cannot be attributed to the different language contexts because we observed the same results for English-speaking subjects when the pitch-shifts were made gradually over successive trials (Jones & Munhall, 2000).

Differences in natural F_0, the range in the duration of phonemes and the manner of speech as well as other factors make it difficult to compare languages even when the data are the result of similar paradigms. However, the similar pattern observed for both English and Mandarin speakers when exposed to the same perturbation paradigms suggests that similar mechanisms underlie the responses.

4. General discussion

The Mandarin-speaking subjects in these experiments showed a similar reliance on F_0 feedback to the English- and Japanese-speaking subjects from previous pitch perturbation experiments. In Experiment 1, the subjects all showed rapid compensatory responses when presented with large, sudden pitch perturbations (Kawahara, 1995; Burnett *et al.*, 1998). In Experiment 2, the subjects showed compensations when small incremental changes in pitch feedback were introduced without their awareness. When vocal pitch feedback was returned to normal, the subjects overcompensated and showed evidence of a negative aftereffect (Jones & Munhall, 2000). These results, in combination, suggest that both postural and phonemic control of vocal pitch involve the use of closed-loop feedback as well as internal representations.

The rationale for the conclusions in this study is based on the status of tone linguistically and experimentally in these studies. The assumption here is that tone is more closely tied to segmental structure than vocal pitch in English is. While vowels in most languages show intrinsic fundamental frequency differences (i.e., F_0 for /i/ is higher than for /æ/), these differences are small biomechanical differences that are not independent of vowel quality (cf. Whalen & Levitt, 1995). In contrast, the vocal pitch contour of vowels in tone languages is largely independent of vowel identity and contrastively determines word meaning locally (i.e., not suprasegmentally). The similarity of the behavior of Mandarin and English subjects indicates that vocal pitch control for vastly different linguistic purposes and for different degrees of mandated precision is accomplished in the same manner.

One possible objection to this conclusion could be that the experimental paradigm removed the linguistic use of tone for our subjects from the planning and control process. By having subjects produce prolonged and repetitive sequences they no longer were producing Mandarin morphemes and thus no longer were using vocal pitch contrastively. By this reasoning, both our English and Mandarin speakers were simply producing a loosely constrained vocal pitch. While we cannot rule out this possibility completely, we think it unlikely for a number of reasons. First, subjects read an ideogram for a particular pronunciation and produced pitch patterns that matched the tone contour for that morpheme (high, flat pitch contour). Second, anecdotally the subjects indicated that they were pronouncing a Mandarin lexical item and talked after the experiment about repeating the same *word* (“mother”) over and over again. Finally, the data themselves support the view that the Mandarin subjects were treating these utterances as tones. The baseline F_0 for the two different sessions in Experiment 2 yielded the same pitch contours with remarkably similar average F_0 values (<2 Hz difference between sessions). This is in contrast to the findings for the English-speaking subjects in Jones & Munhall (2000) who showed variable F_0 baselines across sessions.

Concerns about the linguistic intention of talkers and their control of linguistic variables in speech production studies are not restricted to our paradigm. It is common that studies of speech motor control involve complex articulation but little or no language. Thus, laboratory manipulations of stress, accent, focus, rate, etc., involve significant speech motor control processing but may not involve the natural linguistic use of these variables and thus may not involve natural speech planning processes. This is particularly a concern in studies of the

neural substrates supporting speech and language using functional imaging (Munhall, in press).

The aftereffects observed here and in our previous work (Jones & Munhall, 2000) indicate the presence of an internal representation for vocal pitch. A number of studies of arm movement have shown that subjects exposed to novel conditions modify or acquire new internal models. For example, Shadmehr & Mussa-Ivaldi (1994) found that subjects asked to move a robot manipulandum to targets while the robot-imposed novel forces initially produced distorted trajectories. However, after some practice they made movements similar to those that they made prior to exposure to the artificial force field. When the forces were suddenly removed, subjects showed aftereffects for a few trials and generated torques as though they were still encountering the force field. The results suggest that a new mapping between the kinematics of the arm movements and the forces needed to control trajectories was learned. Subjects seemed to have either modified their internal model and adjusted parameters relating kinematics and muscle forces to adapt to their new force environment, or acquired an entirely new model.

Similar aftereffects are observed when visual information regarding movements is spatially displaced. For example, in a now classic study, Held (1965) had subjects wear prisms that horizontally displaced their visual field. Subjects initially made errors in the direction of the prism displacement when reaching. However, after a relatively short period of practice, reaching speed and accuracy returned to near normal. Similar to observations by Shadmehr & Mussa-Ivaldi (1994), when subjects performed the same reaching task immediately after the prisms were removed, they made errors in the direction opposite to the prism displacement. Thus, a remapping between visual coordinate space and movements occurred.

Our studies suggest that talkers either learn new internal models of laryngeal control or remap the relationship between control parameters and the resultant pitch irrespective of the linguistic use of vocal pitch. In one sense, the evidence for adaptation to modified tone feedback is not surprising. Houde & Jordan (1998) using a similar paradigm showed both compensation and adaptation of the formant frequencies of whispered vowels when feedback was modified.

This finding along with our evidence for compensation and adaptation of F_0 in both English and Mandarin speakers suggests that a wide range of speech parameters are controlled using auditory feedback in similar ways. How then are we to account for the differential stability of phonemic variables and postural variables following changes in hearing status (Cowie & Douglas-Cowie, 1992; Perkell *et al.*, 1997)? Postlingually deafened adults more frequently and rapidly show changes in such postural settings as speaking volume and vocal pitch while consonant and vowel quality are less affected. One possible clue comes from considering the differences in the feedback modification that occurs in deafness vs. our studies. The elimination of auditory feedback with deafness may invoke a different control strategy than when auditory feedback is present, albeit distorted.

In closing, we note that still little is known about the form of the auditory feedback system in speech motor control. The data suggest that the removal of most or all of the auditory feedback induces sudden changes in some speech settings while other aspects of speech (e.g., consonant and vowel characteristics) are robust under these conditions. This finding would suggest that consonants and vowels have

stronger internal representations. However, the compensation and adaptation results from our laboratory and others (e.g., Houde & Jordan, 1998) do not support this dichotomy. Vocal pitch, whether it is involved in controlling the habitual pitch of an English speaker or controlling the tone signaling a morpheme in Mandarin, shows the kinds of compensations that indicate feedback-based control and adaptations consistent with internal models of pitch control.

This work was funded by NIH grant DC-00594 from the National Institute of Deafness and other Communications Disorders and NSERC. We wish to thank Helen Chang and Clare McDuffee for their help in analyzing the data in Experiment 1.

References

Abbs, J. H. & Gracco, V. L. (1984) Control of complex motor gestures: orofacial muscle responses to load perturbations of the lip during speech, *Journal of Neurophysiology*, **51,** 705–723.

Apel, W. (editor). (1969) *Harvard dictionary of music*. Cambridge, MA: Harvard University Press.

Baum, S. & Pell, M. (1999) The neural bases of prosody: insights from lesion studies and neuroimaging, *Aphasiology*, **13,** 581–608.

Binnie, C. A., Daniloff, R. G. & Buckingham H. W. (1982) Phonetic disintegration in a five-year-old following sudden hearing loss, *Journal of Speech and Hearing Disorders,* **47,** 181–189.

Boersma, P. (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, **17,** 97–110.

Burnett, T. A., Senner, J. E. & Larson, C. R. (1997) Voice F_0 responses to pitch-shifted auditory feedback: a preliminary study, *Journal of Voice,* **11,** 202–211.

Burnett, T. A., Freedland, M. B., Larson, C. R. & Hain, T. C. (1998) Voice F_0 responses to manipulations in pitch feedback, *Journal of the Acoustical Society of America*, **103,** 3153–3161.

Clements, G. N. (1985) The geometry of phonological features, *Phonology*, **2,** 225–252.

Clements, G. N. & Hume, E. V. (1995) The internal organization of speech sounds. In *The handbook of phonological theory* (J. A. Goldsmith, editor). Cambridge: Blackwell.

Coleman, R. F. & Markham, I. W. (1991) Normal variations in habitual pitch, *Journal of Voice*, **5,** 173–177.

Cowie, R. & Douglas-Cowie, E. (1992) Postlingually acquired deafness. In *Trends in linguistics, studies and monographs*, Vol. 62. New York: Mouton de Gruyter.

Desmurget, M. & Grafton, S. (2000) Forward modeling allows feedback control for fast reaching movements, *Trends in Cognitive Science*, **4,** 423–431.

Elman, J. L. (1981) Effects of frequency-shifted feedback on the pitch of vocal productions, *Journal of the Acoustical Society of America*, **70,** 45–50.

Flanagan, J. R. & Wing, A. M. (1993) Modulation of grip force with load force during point-to-point arm movements, *Experimental Brain Research*, **95,** 131–143.

Folkins, J. W. & Abbs, J. H. (1975) Lip and jaw motor control during speech: Responses to resistive loading of the jaw, *Journal of Speech and Hearing Research*, **18,** 207–220.

Gandour, J. T. (1978) The perception of tone. In *Tone: a linguistic study* (V. A. Fromkin, editor), New York: Academic Press.

Garber, S. R. & Moller, K. T. (1979) The effects of feedback filtering on nasalization in normal and hypernasal speakers, *Journal of Speech and Hearing Research*, **22,** 321–333.

Goldsmith, J. A. (1976) Autosegmental phonology. Doctoral dissertation, MIT, New York: Garland Press.

Gracco, V. L. & Abbs, J. H. (1988) Central patterning of speech movements, *Experimental Brain Research*, **71,** 515–526.

Gramming, P., Sundberg, J., Ternström, L. & Perkins, W. H. (1988) Relationship between changes in voice pitch and loudness, *Journal of Voice*, **2,** 118–126.

Guthrie, B. L., Porter, J. D. & Sparks, D. L. (1983) Corollary discharge provides accurate eye position information to the oculomotor system, *Science*, **221,** 1193–1195.

Hamlet, S. & Stone, M. (1976) Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses, *Journal of Phonetics*, **4,** 199–218.

Hamlet, S. & Stone, M. (1978) Compensatory alveolar consonant production induced by wearing a dental prosthesis, *Journal of Phonetics*, **6,** 227–248.

Hamlet, S., Stone, M. & McCarty, T. (1978) Conditioning dentures viewed from the standpoint of speech adaptation, *Journal of Prosthetic Dentistry*, **40,** 60–66.

Held, R. (1965) Plasticity in sensory-motor systems. *Scientific American*, **213,** 84–94.

Honda, M. & Fujino, A. (2000). Articulatory compensation and adaptation for unexpected palate shape perturbation. *Proceedings of the 6th International Conference on Spoken Language Processing*. Beijing.

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

Johansson, R. S. & Westling, G. (1984) Roles of glabrous skin receptors and sensorimotor memory in automatic-control of precision grip when lifting rougher or more slippery objects, *Experimental Brain Research*, **56,** 550–564.

Jones, J. A. & Munhall, K. G. (2000) Perceptual calibration of F_0 production: evidence from feedback perturbation, *Journal of the Acoustical Society of America*, **108,** 1246–1251.

Kawahara, H. (1995) Hearing voice: transformed auditory feedback effects on voice pitch control. *Proceedings of the international joint conference on artificial intelligence: workshop on computational auditory scene analysis*, pp. 143–148. Montreal, Canada

Keller, E. L., Gandhi, N. J. & Shieh, J. M. (1996) Endpoint accuracy in saccades interrupted by stimulation in the ominipause region in monkey, *Visual Neuroscience*, **13,** 1059–1067.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E. & Fowler, C. (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures, *Journal of Experimental Psychology: Human Perception and Performance*, **10,** 812–832.

Kollia, H. B., Gracco,V. L. & Harris, K. S. (1992) Functional organization of velar movements following jaw perturbation, *Journal of the Acoustical Society of America*, **91,** 2474.

Kratochvil, P. (1998) Intonation in Beijing Chinese. In *Intonation systems* (D. Hirst & A. Di Christo, editors). Cambridge: Cambridge University Press.

Lane, H. & Tranel, B. (1971) The Lombard sign and the role of hearing in speech, *Journal of Speech and Hearing Research*, **14,** 677–709.

Lane, H. & Webster, J. W. (1991) Speech deterioration in postlingually deafened adults, *Journal of the Acoustical Society of America*, **89,** 859–866.

Larson, C. R. (1998) Cross-modality influences in speech motor control: the use of pitch shifting for the study of F_0 control, *Journal of Communication Disorders*, **31,** 498–503.

Larson, C. R., Burnett, T. A., Kiran, S. & Hain, T. C. (2000) Effects of pitch-shift velocity on voice F_0 responses, *Journal of the Acoustical Society of America*, **107,** 559–564.

Löfqvist, A. & Gracco, V. L. (1991) Discrete and continuous modes in speech motor control, *PERILUS*, XIV, 27–34.

McFarland, D. H., Baum, S. R. & Chabot, C. (1996) Speech compensation to structural modifications of the oral cavity, *Journal of the Acoustical Society of America*, **100,** 1093–1104.

Munhall, K. G. (in press) Functional imaging during speech production. *Acta Psychologica*.

Munhall, K., Löfqvist, A. & Kelso, J. A. S. (1994) Lip-larynx coordination in speech: effects of mechanical perturbations to the lower lip, *Journal of the Acoustical Society of America*, **96,** 3605–3616.

Odden, D. (1995) Tone: African languages. In *The handbook of phonological theory* (J. A. Goldsmith, editor). Cambridge: Blackwell.

Oller, D. & Eilers, R. (1988) The role of audition in infant babbling, *Child Development*, **59,** 441–449.

Pelisson, D., Guitton, D. & Goffart, L. (1995) Online compensation of gaze shifts perturbed by microstimulation of the superior colliculus in the cat with unrestrained head, *Experimental Brain Research*, **106,** 196–204.

Perkell, J. S., Matthies, M. L., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J. & Guiod, P. (1997) Speech motor control: acoustic segmental goals, saturation effects, auditory feedback and internal models, *Speech Communication*, **22,** 227–250.

Ringel, R. L. & Steer, M. (1963) Some effects of tactile and auditory alterations on speech output. *Journal of Speech and Hearing Research*, **6,** 369–378.

Shaiman, S. (1989) Kinematic and electromyographic responses to perturbation of the jaw, *Journal of the Acoustical Society of America*, **86,** 78–87.

Shadmehr, R. & Mussa-Ivaldi, F. A. (1994) Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, **14,** 3208–3224.

Smith, C. (1975) Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research,* **18,** 795–811.

Smith, K. U. (1962) *Delayed sensory feedback and behavior*. Philadelphia: W.B. Saunders.

Sorokin, V., Olshansky, V. & Kozhanov, L. (1998). Internal model in articulatory control: evidence from speaking without larynx, *Speech Communication*, **25,** 249–268.

Svirsky, M. A., Lane, H., Perkell, J. S. & Wozniak, J. (1992) Effects of short-term auditory deprivation on speech production in adult cochlear implant users, *Journal of the Acoustical Society of America*, **92,** 1284–1300.

Titze, I. R. (1994) *Principles of voice production*. Englewood Cliffs, NJ: Prentice-Hall.

Waldstein, R. (1990) Effects of postlingual deafness on speech production: implications for the role of auditory feedback, *Journal of the Acoustical Society of America*, **88,** 2099–2114.

Whalen, D. H. & Levitt, A. G. (1995) The universality of intrinsic F_0 of vowels, *Journal of Phonetics*, **23,** 349–366.
Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995) An internal model for sensorimotor integration, *Science*, **269,** 1880–1882.
Wolpert, D. M. & Kawato, M. (1998) Multiple paired forward and inverse models for motor control, *Neural Networks*, **11,** 1317–1329.
Wu, Z. (2000) From traditional Chinese phonology to modern speech processing: realization of tone and intonation in standard Chinese. *Proceedings of the 6th international conference on spoken language processing*. Beijing, China.
Xu, Y. and Sun, X. (2000) How fast can we really change pitch? Maximum speech of pitch change revisited. *Proceedings of the 6th International Conference on Spoken Language Processing*. Beijing, China.
Yip, M. (1995) Tone in east Asian languages. In *The handbook of phonological theory* (J. A. Goldsmith, editor). Cambridge: Blackwell.
Zemlin, W. R. (1981) *Speech and hearing science: anatomy and physiology*, 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.