Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations

Hanjun Liu^{a)}

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

Emily Q. Wang

Department of Communication Disorders and Sciences, Rush University Medical Center, 1653 West Congress Parkway, 203 Senn, Chicago, Illinois 60612

Zhaocong Chen and Peng Liu

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

Charles R. Larson

Department of Communication Science and Disorders, Northwestern University, Evanston, Illinois 60208

Dongfeng Huang

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

(Received 13 February 2010; revised 20 July 2010; accepted 15 September 2010)

The purpose of this cross-language study was to examine whether the online control of voice fundamental frequency (F_0) during vowel phonation is influenced by language experience. Native speakers of Cantonese and Mandarin, both tonal languages spoken in China, participated in the experiments. Subjects were asked to vocalize a vowel sound /u/ at their comfortable habitual F_0 , during which their voice pitch was unexpectedly shifted (± 50 , ± 100 , ± 200 , or ± 500 cents, 200 ms duration) and fed back instantaneously to them over headphones. The results showed that Cantonese speakers produced significantly smaller responses than Mandarin speakers when the stimulus magnitude varied from 200 to 500 cents. Further, response magnitudes decreased along with the increase in stimulus magnitude in Cantonese speakers, which was not observed in Mandarin speakers. These findings suggest that online control of voice F_0 during vocalization is sensitive to language experience. Further, systematic modulations of vocal responses across stimulus magnitude were observed in Cantonese speakers but not in Mandarin speakers, which indicates that this highly automatic feedback mechanism is sensitive to the specific tonal system of each language.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3500675]

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

Pages: 3739–3746

I. INTRODUCTION

Speech production is one of the most complex of human behaviors requiring the integration of auditory and somatosensory feedback with language processes in order to insure accurate and clear communication with others. Auditory feedback is important for learning the relationship between articulation and sound production and is also used to correct errors during production (Guenther, 2006). The fact that the normal development of speech production is severely affected or prevented for those children with prelingual deafness (Oller and Eilers, 1988) indicates that auditory feedback plays a critical role during the acquisition of language. Further, auditory feedback in speech production after language acquisition may still serve an important purpose,

as studies on postlingually deafened adults show degradation in phonatory, articulatory, and suprasegmental control during speech production (Binnie *et al.*, 1982; Lane and Webster, 1991).

Recent research has demonstrated the importance of audition for the control of vocalization by showing that unexpected perturbations in auditory feedback result in changes in voice fundamental frequency (F_0) (Elman, 1981; Kawahara, 1994; Burnett *et al.*, 1998). In this experimental paradigm, participants vocalize a vowel sound or speak a phrase while they hear their voice played back through headphones. During speaking, their voice pitch is unexpectedly shifted up or down. This shift in pitch represents a mismatch between what the subjects intended to say and what was heard. The subjects then attempt to correct for this mismatch by changing their voice F_0 in the direction opposite to the stimulus. Namely, they lower their voice F_0 when their voice pitch feedback is shifted upward and vice versa. It has been

^{a)}Author to whom correspondence should be addressed. Electronic mail: lhanjun@mail.sysu.edu.cn

suggested that this compensatory response is reflexive because the latency is between 100 and 150 ms after the onset of pitch feedback perturbation, and the response is made in the opposite direction to the stimulus; therefore, it is termed “pitch-shift reflex” (PSR). Such compensatory mechanisms of voice control have been widely reported in English-speaking adults (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 2000; Chen *et al.*, 2007; Larson *et al.*, 2007; Liu and Larson, 2007; Jones and Keough, 2008; Hawco *et al.*, 2009). Similar compensatory responses were also observed in speakers of other languages, such as Mandarin (Jones and Munhall, 2002; Xu *et al.*, 2004; Liu *et al.*, 2009) and German (Natke and Kalveram, 2001; Donath *et al.*, 2002; Natke *et al.*, 2003).

Although compensatory responses of voice F_0 control have been observed across different languages, and many interesting observations about the PSR have been obtained from previously published studies, an important question remains unanswered, i.e., whether the modulations of these responses across different stimuli or vocal task are universal across languages or constrained by speakers’ language experience. For example, German speakers produced larger PSR magnitudes during singing than during speaking (Natke *et al.*, 2003). Vocal responses to pitch perturbations that were presented later relative to the vocal onset were significantly reduced as compared to stimuli that were presented earlier during the production of the Mandarin phrase “mao mi mo ma” in a question pattern (Liu *et al.*, 2009). PSR magnitudes were larger during production of a sentence “you know nina?” than during a sustained vowel in English (Chen *et al.*, 2007). However, since the subjects in these studies spoke their own language, English, Mandarin, or German, it is difficult to make cross-language comparisons.

The present study was thus designed to assess the effect of language background by examining vocal responses to pitch feedback perturbations during the same phonation task in Cantonese- and Mandarin-speaking adults. Both Cantonese and Mandarin are considered tonal languages. They differ from non-tonal languages such as English in that the tones in these languages change the meanings of otherwise identical syllables. For example, the syllable “ma” in Mandarin produced with a high level tone means “mother” while “ma” produced with a high falling tone means “to scold.” The primary difference between the two tonal languages is that while Cantonese has six contrastive tones (high level, high rising, mid level, low falling, low rising, and low level), Mandarin, in contrast, has four contrastive tones (high level, high rising, falling rising, and high falling). For male speakers, typically, the six Cantonese tones cover a F_0 range of 80–170 Hz (Francis *et al.*, 2008) while four Mandarin tones cover a F_0 range of 90–140 Hz (Xu, 1997). Therefore, the Cantonese speakers need to be able to differentiate six tones across a F_0 range that is similar to the Mandarin range. The narrower range across which Cantonese speakers must perceive tonal differences may require more highly tuned perceptual abilities than in Mandarin speakers.

Native speakers of a tonal language are superior to speakers of a non-tonal language in perceiving and producing rapid changes in voice F_0 (Eady, 1982). It has also been

demonstrated that pitch perception varies depending on a listener’s familiarity with the pitch patterns that occur in his native language. Tonal language listeners such as Mandarin and Cantonese are better at discriminating the direction of movement of F_0 (e.g., rising or falling) than English listeners (Gandour and Harshman, 1978; Gandour, 1983). However, when compared to Mandarin listeners, Cantonese listeners placed more emphasis on the average F_0 level (i.e., whether the height of F_0 is high or low) (Vance, 1976, 1977; Tse, 1978). In addition, Cantonese listeners have been found to be better at discriminating Mandarin tones than Mandarin listeners are at discriminating Cantonese tones (Lee *et al.*, 1996).

On the production side, recent pitch-shift studies suggest that the ability of a speaker to regulate voice F_0 during speech may be related to the specific tonal system of a language. For example, when Mandarin speakers produced disyllabic sequences “ma ma” with high rising and high falling bi-tonal patterns, greater pitch-shift response magnitudes were observed for downward than for upward stimuli (Xu *et al.*, 2004). This sensitivity of the vocal responses to stimulus direction was not observed when native Mandarin speakers produced all the syllables of a sentence with high level tones (Liu *et al.*, 2009). Thus, there seems to be a differential effect of the pattern of tonal modulations in Mandarin on the modulation of vocal responses to pitch feedback perturbations. It is noted that Mandarin and Cantonese differ not only in the number of tones (six vs four) but also that Cantonese speakers place more emphasis on the perception of the height of level tones than do Mandarin speakers (Vance, 1976, 1977; Tse, 1978), suggesting that Cantonese speakers may monitor their voice F_0 more tightly to produce correct tonal patterns during sustained vocalization than Mandarin speakers. In this case, Cantonese speakers may attenuate their vocal responses to the perturbations in the auditory feedback during the production of sustained vowels, which is analogous to the observation that musicians produced small or no compensatory responses to pitch feedback perturbations when they were asked to hit a specific note (Zarate and Zatorre, 2008). It is thus argued that the difference in the tonal patterns between Mandarin and Cantonese may lead to a differential processing of pitch feedback perturbations, suggesting that online control of voice F_0 may be specific within different tonal systems. The exploration of these mechanisms may lead to further understanding of the relationship between language experience and neural plasticity underlying the audio-vocal system.

In this cross-language study, therefore, we hypothesized that Cantonese speakers would produce smaller vocal responses to pitch feedback perturbations than Mandarin speakers during sustained vocalization, which would relate to the specific features of the native tonal system of each tonal language. Moreover, since previous English studies showed effects of stimulus magnitude on the vocal responses to small pitch perturbations (e.g., 50 cents) but not to large ones (e.g., 200 cents) (Chen *et al.*, 2007; Liu and Larson, 2007), we also tested the effects of stimulus direction (up and down) and stimulus magnitude (50, 100, 200, and 500 cents) on responses to pitch perturbations.

II. METHODS

A. Subjects

Eighteen native Cantonese subjects (nine males and nine females; aged 19–27) and 18 native Mandarin subjects (nine males and nine females; aged 19–27), who were students from Sun Yat-sen University of China, participated in the experiments. Cantonese speakers spoke both Cantonese and Mandarin, but Cantonese was the language they spoke in daily life. Mandarin speakers spoke Mandarin only and none of them were capable of speaking or understanding Cantonese. All subjects reported no history of any hearing, language, speech, or neurological disorders. All subjects signed the consent form approved by Institutional Review Board of The First Affiliated Hospital at Sun Yat-sen University of China.

B. Apparatus

Subjects were seated comfortably in a sound booth throughout the testing. The participants' voices were recorded through a Genuine Shupu microphone (model SM-306) with a mouth-to-microphone distance of 2.5 cm. The microphone signal was amplified with a MOTU Ultralite Mk3 firewire audio interface, shifted in pitch with an Eventide Eclipse Harmonizer, and then played back to subjects through headphones (model T20RP mkII). MIDI software (Max/MSP, v.5.0 by Cycling 74) was used to control the harmonizer. Acoustic calibration was done to insure that the sound level of the artificial feedback was 10 dB (sound pressure level, SPL) higher than that of the natural unamplified airborne feedback. The voice output, feedback, and transistor-transistor logical (TTL) control pulses indicating the onset and offset of the stimulus were digitized at 10 kHz by PowerLab (model ML880, AD Instruments, Castle Hill, Australia) and recorded using LABCHART software (v.7.0 by AD Instruments).

C. Procedure

During the experiments, subjects were asked to vocalize a vowel sound /u/ for approximately 5 s at their comfortable habitual F_0 , during which their voice pitch feedback was shifted either upward or downward five times (random inter-stimulus interval 700–900 ms) and fed back instantaneously to them through headphones. Each block of trials consisted of 12 consecutive vocalizations, and in each block, magnitudes of pitch feedback perturbations were fixed at ± 50 , ± 100 , ± 200 , or ± 500 cents (100 cents = one semitone) with a duration of 200 ms, leading to 30 upward and 30 downward perturbations for each condition. During each vocalization, pitch-shifted stimuli were presented with a delay of 500–700 ms after vocal onset.

D. Response measurement

For the measurement of vocal responses to pitch feedback perturbations, the voice signal was processed by PRAAT (Boersma, 2001) using an autocorrelation method to produce a train of pulses corresponding to the fundamental period of the voice waveform. This pulse train was transformed to an

F_0 contour wave in IGOR PRO (v.6.0 by Wavemetrics, Inc., Lake Oswego, OR) and then converted to a cent scale using the following:

$$\text{Pitch (cents)} = 100 \times (12 \times \log(f_2/f_1)/\log(2))$$

where f_1 is an arbitrary reference note at 195.997 Hz (G4), and f_2 is the voice signal in hertz.

Event-related averaging techniques were used to measure the magnitude and latency of vocal responses (Chen *et al.*, 2007; Larson *et al.*, 2008). Specifically, voice signals in cents were time-aligned with the TTL pulse, using a window of 200 ms pre- and 700 ms post-stimulus. Prior to the averaging, visual inspections were performed on each individual trial so that those trials with unusually large amplitude resulting from either signal processing errors or vocal interruption were removed from further analysis. The criterion for defining a response was a change in the F_0 contour that exceeded a value of two standard deviations (SDs) of the pre-stimulus mean (baseline F_0) beginning at least 60 ms after the stimulus and lasting at least 50 ms (Chen *et al.*, 2007). The criterion for the end of a response was that it returned to within two SDs of the pre-stimulus mean for at least 30 ms. Latency of the averaged response was measured as the time from the stimulus onset at which the response exceeded two SDs of the pre-stimulus mean, and the response magnitude was measured as the difference between the pre-stimulus mean and the greatest or lowest value of the F_0 contour following the response onset (see Fig. 1).

FIG. 1. (A) Contours of voice F_0 (top trace) and auditory feedback (bottom trace). The arrows in the contour of auditory feedback denote the pitch-shift perturbations (± 100 cents) presented during vocalization. (B) One single averaged vocal response to 100 cents stimuli. Horizontal dense dotted lines represent ± 2 SDs of the pre-stimulus mean averaged F_0 . Vertical dash lines indicate the onset and offset times of the responses. Horizontal sparse dash lines indicate the place where response magnitude is measured. Square bracket denotes the direction and time of the pitch-shift stimulus at bottom.

TABLE I. The number of OPP, FOL, and NR as a function of stimulus direction and stimulus magnitude for Cantonese speakers.

		OPP	FOL	NR	Total
Stimulus magnitude	50 cents	32	2	2	36
	100 cents	30	3	3	36
	200 cents	32	4	0	36
	500 cents	22	11	3	36
Stimulus direction	Up	53	12	7	72
	Down	63	8	1	72

E. Statistical analysis

Values of response magnitude and latency were submitted to significance testing in SPSS (v.16.0). Prior to the statistical analysis, tests of normal distribution (Kolmogorov–Smirnov test) and homogeneity of variance (Levene’s test) were performed on the response magnitudes and latencies to verify that the assumptions of analysis of variance (ANOVA) were satisfied. It is noted that these data lend themselves to repeated-measures ANOVA (Xu *et al.*, 2004); however, factorial ANOVAs were used in those cases when the assumptions of repeated-measures ANOVA were not met because of missing data and unequal cell size (Bauer *et al.*, 2006). In those cases where the assumptions of an ANOVA were not met, a non-parametric test (Kruskal–Wallis test) was used for significance testing.

III. RESULTS

144 responses (18 subjects $\times$ 4 stimulus magnitudes $\times$ 2 stimulus directions) were obtained from each of the two language groups: Mandarin and Cantonese. The responses were separated into three types: “following” (FOL), opposing (OPP), and non-response (NR), and the number of these three responses are shown in Tables I and II across stimulus direction and stimulus magnitude for Cantonese and Mandarin speakers. For Cantonese speakers, it can be seen from Table I that 20 of 144 responses “followed” the direction of the stimulus, and the majority of the FOL responses (11 of 20) occurred under the 500 cents condition. Eight of 144 responses did not meet the criteria of either an OPP or FOL response and were defined as NRs, in which 7 of 8 NRs occurred to the upward stimuli. In contrast, for the Mandarin speakers, only 11 of 144 responses followed the direction of the stimulus, and there were three NRs (Table II). Similar to

TABLE II. The number of OPP, FOL, and NR as a function of stimulus direction and stimulus magnitude for Mandarin speakers.

		OPP	FOL	NR	Total
Stimulus magnitude	50 cents	36	0	0	36
	100 cents	36	0	0	36
	200 cents	33	2	1	36
	500 cents	25	9	2	36
Stimulus direction	Up	65	5	2	72
	Down	65	6	1	72

FIG. 2. Representative averaged F_0 contours to pitch perturbations generated by Cantonese speakers as a function of stimulus direction and stimulus magnitude. Left and right columns show the vocal responses to upward and downward stimuli, respectively. From top to bottom are shown the vocal responses to 50, 100, 200, and 500 cents stimuli.

Cantonese speakers, most of the FOL responses (9 of 11) occurred in the 500 cents condition.

Figures 2 and 3 show the representative averaged F_0 contours to pitch feedback perturbations as a function of stimulus direction and stimulus magnitude for Cantonese and Mandarin speakers, respectively. Averaged F_0 contours to upward and downward stimuli are shown in the left and the right columns, and averaged F_0 contours to 50, 100, 200, and 500 cents stimuli are displayed from the top to the bottom. It can be observed that response magnitudes to 200 and 500 cents stimuli are smaller than those to 50 and 100 cents stimuli for the Cantonese speaker (Fig. 2), while the Mandarin speaker produced similar response magnitudes across all stimulus magnitudes (Fig. 3). Response magnitudes to upward stimuli (left) did not differ from those to downward stimuli (right) for either the Cantonese or Mandarin speakers as shown in Figs. 2 and 3.

A. Response magnitude latency for Cantonese and Mandarin

A three-way (language group $\times$ stimulus direction $\times$ stimulus magnitude) factorial ANOVA without repeated measures was used to test for response magnitude. The results showed a main effect of language group [$F(1, 230) = 11.536, p = 0.001$], where Mandarin speakers produced significantly larger response magnitudes (14.1 ± 7.2 cents)

FIG. 3. Representative averaged F_0 contours to pitch perturbations generated by Mandarin speakers as a function of stimulus direction and stimulus magnitude. Left and right columns show the vocal responses to upward and downward stimuli, respectively. From top to bottom are shown the vocal responses to 50, 100, 200, and 500 cents stimuli.

than Cantonese speakers (11.7 ± 5.7 cents). No significant effects of the stimulus direction [$F(1, 230) = 3.390$, $p = 0.067$] or stimulus magnitude [$F(3, 230) = 1.796$, $p = 0.149$] were found on the response magnitudes. A significant interaction was found between language group and stimulus magnitude [$F(3, 230) = 2.512$, $p = 0.04$]. *Post-hoc* paired t -tests showed that Cantonese speakers produced significantly smaller vocal responses than Mandarin speakers to 200 cents (9.8 vs 14.7 cents) ($t = 2.563$, $df = 21$, $p = 0.018$) and 500 cents stimuli (9.0 vs 14.4 cents) ($t = 2.880$, $df = 28$, $p = 0.008$) (see Fig. 4). However, both Cantonese and Mandarin produced similar vocal responses to 50 cents (12.9 vs 13.0 cents) ($t = 0.042$, $df = 31$, $p = 0.967$) and 100 cents stimuli (13.4 vs 14.6 cents) ($t = 0.850$, $df = 32$, $p = 0.402$).

Since the response latency data did not have a normal distribution (Kolmogorov–Smirnov test: $Z = 3.165$, $p < 0.001$) and did not pass the test of homogeneity of variance [Levene’s test: $F(15, 230) = 4.848$, $p < 0.001$], non-parametric Kruskal–Wallis tests were used to test the significance of response latency across the conditions. Statistical results revealed no significant differences for response latencies in the language group ($\chi^2 = 0.326$, $df = 1$, $p = 0.568$) or stimulus magnitude ($\chi^2 = 4.733$, $df = 3$, $p = 0.192$) (see Fig. 5). Significantly longer latencies, however, were found for upward stimuli (116 ± 76 ms) as compared to downward stimuli (96 ± 55 ms) ($\chi^2 = 7.604$, $df = 1$, $p = 0.006$).

FIG. 4. Box plots illustrating the response magnitude as a function of stimulus magnitude for Cantonese and Mandarin speakers. Box plot definitions: Middle line is median, top and bottom of boxes are 75th and 25th percentiles, and whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge – low hinge) and low hinge – 1.5 (high hinge – low hinge).

Due to the significant interaction between language group and stimulus magnitude, separate statistical tests were performed to explore the significant effects of stimulus direction and magnitude on response magnitude and latency for Cantonese and Mandarin speakers, respectively. Prior to the statistical analysis, tests of normal distribution and homogeneity of variance revealed that response latency data produced by Cantonese speakers failed to meet the assumptions of parametric ANOVAs [Kolmogorov–Smirnov test: $Z = 2.671$, $p < 0.001$; Levene’s test: $F(7, 108) = 7.663$, $p < 0.001$], so a non-parametric Kruskal–Wallis test was used for significance testing of the responses’ latencies. Although response magnitude data of Cantonese speakers met these assumptions of parametric tests, a factorial ANOVA without repeated measures was used because of missing data and unequal cell size (see Table I). Regarding the data of response magnitude and latency produced by Mandarin speakers, repeated-measured ANOVAs were used for significance tests across stimulus direction and magnitude.

B. Response magnitude and latency for Cantonese

In Cantonese speakers, a factorial ANOVA revealed significant differences in response magnitudes as a function of

FIG. 5. Box plots illustrating the response latency as a function of stimulus magnitude for Cantonese and Mandarin speakers.

stimulus magnitude [$F(3, 108) = 4.460, p = 0.005$]. A *post-hoc* least significant difference (LSD) test showed that 200 and 500 cents stimuli generated significantly smaller response magnitudes (9.8 ± 3.7 ; 9.0 ± 4.3 cents, respectively) than 50 (12.9 ± 5.8 cents) and 100 cents stimuli (13.4 ± 7.5 cents) ($p < 0.03$) (see Fig. 4). Vocal responses to upward stimuli (10.6 ± 4.7 cents) did not differ significantly from those to downward stimuli (12.3 ± 6.7 cents) [$F(1, 108) = 3.260, p = 0.074$]. No significant interactions between stimulus direction and stimulus magnitude [$F(3, 108) = 1.313, p = 0.274$] were observed. Regarding the response latency of Cantonese speakers, a Kruskal–Wallis test revealed no significant differences for stimulus magnitude ($\chi^2 = 5.528, df = 3, p = 0.137$) or stimulus direction ($\chi^2 = 1.375, df = 1, p = 0.241$).

C. Response magnitude and latency for Mandarin

In Mandarin speakers, a repeated-measures ANOVA for response magnitude revealed no main effects of stimulus magnitude [$F(3, 51) = 0.694, p = 0.560$] or stimulus direction [$F(3, 51) = 2.369, p = 0.142$] as well as stimulus magnitude $\times$ stimulus direction interaction [$F(3, 51) = 1.459, p = 0.237$]. Similarly, statistical results for response latency showed no main effects of stimulus magnitude [$F(3, 51) = 1.704, p = 0.178$] and stimulus direction [$F(1, 17) = 4.309, p = 0.053$] or significant interactions with stimulus direction [$F(3, 51) = 0.710, p = 0.550$].

IV. DISCUSSION

This cross-language study investigated whether the online processing of auditory feedback during vocalization is sensitive to language experience by examining the vocal responses to F_0 perturbations produced by speakers of Cantonese and Mandarin. The present findings showed that Cantonese speakers, as we hypothesized, produced smaller response magnitudes than Mandarin speakers for large stimulus magnitudes (200 and 500 cents). Further, Cantonese, but not Mandarin speakers, also produced smaller response magnitudes for the large stimuli (200 and 500 cents) compared to the small stimuli (50 and 100 cents). These findings suggest that online monitoring of voice F_0 may differ across different tonal languages, and this difference may be accounted for by the specific tonal system of each language.

A. Cantonese vs Mandarin

As discussed in Sec. I, Cantonese speakers are better at discriminating Mandarin tones than Mandarin speakers at discriminating Cantonese tones (Lee *et al.*, 1996). The relative ease with which Cantonese speakers can discriminate Mandarin tones may be due to the complexity of the Cantonese tonal system. Cantonese has three level tones (high level, middle level, and low level) and two rising tones (high rising and low rising), and these two dynamic tones differ only in the onset frequency but not in their tonal patterns. In contrast, Mandarin has only one static tone (high level), which would be easier to differentiate from the other three dynamic tones: One with a rising pattern, one with a falling pattern, and the third with a falling–rising pattern. Further, develop-

mental studies on tone acquisition suggest that Cantonese tones are more difficult to acquire than Mandarin tones: Mandarin children learn all four Mandarin tones by stage 2 (17–20 months) of child language development, while Cantonese children still have two more Cantonese tones to acquire at the same stage (Li and Thompson, 1977; Tse, 1978). These studies show that Cantonese not only has a more complex tonal system than Mandarin, which may require a longer time to acquire, but it may also affect the sensitivity and discriminability of Cantonese speakers to different tonal patterns.

This sensitivity to the level tones that Cantonese speakers have may account for the findings that Cantonese speakers produced smaller response magnitudes than Mandarin speakers for large stimulus magnitudes (i.e., 200 and 500 cents). Cantonese tones have three level tones, and the F_0 height differences among these three level tones vary from 10 to 40 Hz (Francis *et al.*, 2008). In the present study, when pitch feedback perturbation varied from 50, 100, 200, to 500 cents, the corresponding shifts were 4.0, 8.9, 18.3, and 50.0 Hz if voice F_0 was maintained at 150 Hz. Therefore, pitch feedback perturbations of 50 or 100 cents (4.0 and 8.9 Hz) were still within the F_0 range for Cantonese level tones (10 Hz). When 200 or 500 cents stimuli were introduced, however, the magnitude of the perturbations (18.3 or 50 Hz) may have exceeded the F_0 height differences between the two level tones in Cantonese, leading to a misperception of another level tone. By contrast, Mandarin has one single level tone (high level), and the large perturbations may have been perceived as a level tone within the range of a single tone rather than a tone of a different level. This difference between the level tones in Cantonese and Mandarin suggest that Cantonese speakers may have developed different strategies for the differential processing of small and large pitch feedback perturbation in voice F_0 control, which is related to the tonal patterns specific to this language.

Specifically, Cantonese speakers may have learned to use auditory feedback to help control the tone level within a narrow range of frequencies that is specific for each of the three tone levels. Thus, their responses to the 50 and 100 cents stimuli were similar to those of the Mandarin speakers because the perturbations were perceived to be within a single tone. However, when large perturbations (200 and 500 cents) were presented, responses of Cantonese speakers were smaller than Mandarin responses because the Cantonese speakers did not perceive the stimuli to be within the range of a single tone. This suggestion is based on the idea that speakers of a tonal language learn the frequency range of a specific tone (e.g., high level) and they rely on automatic mechanisms of error correction for F_0 variations within this range. Beyond this range of frequencies, they learn that large perturbations are indicative of a completely different level and hence do not require a corrective response as large as the one produced within the frequency range of a single tone.

B. English vs Chinese (Mandarin and Cantonese)

Since the effects of stimulus magnitude on the PSR during vowel phonation have been extensively examined in

previous English pitch-shift studies (Burnett *et al.*, 1998; Chen *et al.*, 2007), it is thus possible to compare those studies to our results. In the present study, Mandarin speakers produced overall response magnitudes of 14.1 cents and latencies of 105 ms, while for Cantonese speakers they were 11.7 cents and 103 ms when stimulus magnitudes varied from 50, 100, 200, to 500 cents. In contrast, Chen *et al.* (2007) reported response magnitudes and latencies of 21.6 cents and 154 ms for English speakers in responding to pitch-shift stimuli of 50, 100, and 200 cents during sustained vowels. In a study by Burnett *et al.* (1998), the overall response magnitudes were 29.2 cents and response latencies were 266 ms. Larson *et al.* (2001) varied stimulus magnitudes from 25, 100, to 200 cents, and the resultant vocal response magnitudes were 26, 48, and 39 cents, respectively, and latencies of 246, 117, and 130 ms, respectively. In several other English pitch-shift studies, however, similar vocal responses were produced compared to the present study. For example, Burnett *et al.* (2008) reported a mean response of 11.25 cents and 115 ms to 100 cents stimuli with 100 ms duration. Response to this stimulus magnitude was reported as 14.3 cents and 135 ms by Bauer and Larson (2003). Despite the minor methodological differences between these English studies and the present study, the findings altogether suggest that, as compared to English speakers, both Mandarin and Cantonese speakers may produce faster vocal responses to pitch feedback perturbations. Whether there is a difference in response magnitude between English and Chinese, however, seems to be controversial and needs to be clarified in further experiments.

The shorter response latencies observed in this study in Chinese speakers when compared to those in English speakers indicate that speakers of tonal languages may process pitch feedback perturbations faster than speakers of non-tonal language. Such differences in the response latency were also observed during speech production. The overall response latency to pitch feedback perturbation was 122 ms (Chen *et al.*, 2007) and 133 ms (Liu *et al.*, 2010) during the production of “you know nina?” by English speakers. However, the overall response latency was 109 ms when Mandarin speakers spoke a Mandarin sentence of “mao mi mo ma?” (Liu *et al.*, 2009). Previous research has demonstrated that speakers of a tonal language may have an advantage in processing the rapid pitch changes because in any given time interval, speakers of tonal languages such as Mandarin and Cantonese have to produce more dynamic F_0 movements and process more syllables than speakers of non-tonal languages such as English (Eady, 1982). Therefore, the required ability for rapid processing of F_0 changes in Mandarin and Cantonese may also lead the speakers of these tonal languages to respond faster to pitch feedback perturbations during vowel phonation.

C. Implications for future research

Along with previous English studies, this cross-language study suggests that vocal responses to pitch feedback perturbations are dependent on the intrinsic linguistic system of each language. A person’s linguistic experience can shape the audio-vocal system that constrains his/her automatic/reflexive responses to stimuli such as pitch perturbations.

The findings that the online control of voice F_0 is sensitive to a speaker’s long-term experience with his/her native language suggest that the audio-vocal system is amenable to changes as required by the specific language. This neural plasticity indicates there is an intimate link between speech production and speech perception and helps us understand how speech perception interacts with speech production. Although neural plasticity in encoding linguistic pitch patterns has been observed at the cortical and subcortical level after long- or short-term speech learning or music training (Wong *et al.*, 2007; Song *et al.*, 2008; Atiani *et al.*, 2009), whether such training can also induce the neural plasticity of the audio-vocal system in the processing of auditory feedback is largely unknown. Future studies should be conducted to explore how mechanisms of voice F_0 control can change when a speaker’s perception of tonal changes is improved after training.

V. CONCLUSION

The present study examined whether online control of voice F_0 is language experience specific by comparing vocal responses to pitch feedback perturbations generated by Cantonese speakers with those by Mandarin speakers. The results showed that Cantonese speakers differed from Mandarin speakers in the online monitoring of auditory feedback during vocalization. Larger magnitudes of vocal responses were associated with Mandarin speakers than Cantonese speakers at large pitch-shift stimuli. Cantonese speakers, but not Mandarin speakers, decreased their vocal responses with the increase in stimulus magnitude. These findings suggest that the online control of voice F_0 is sensitive to language experience. Systematic modulations of vocal responses across stimulus magnitude in Cantonese speakers but not in Mandarin speakers indicate that this automatic compensation for voice pitch errors is constrained by the specific tonal system of each language.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China (Grant No. 30970965), Guangdong Natural Science Foundation (Grant No. 9151008901000053), and NIH Grant No. 1R01DC006243. The authors thank Dr. Yi Xu at University of College London for his helpful discussion regarding Mandarin tonal production. Thanks also to Chun Liang Chan at Northwestern University for his programming help.

- Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). “Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields,” *Neuron* **61**, 467–480.
- 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.

- Binnie, C. A., Daniloff, R. G., and Buckingham, H. W. (1982). "Phonetic disintegration in a five-year-old following sudden hearing loss," *J. Speech Hear. Disord.* **47**, 181–189.
- Boersma, P. (2001). "Praat, a system for doing phonetics by computer," *Glott Int.* **5**(9/10), 341–345.
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F_0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am.* **103**, 3153–3161.
- Burnett, T. A., McCurdy, K. E., and Bright, J. C. (2008). "Reflexive and volitional voice fundamental frequency responses to an anticipated feedback pitch error," *Exp. Brain Res.* **191**, 341–351.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice F_0 responses to pitch-shifted voice feedback during English speech," *J. Acoust. Soc. Am.* **121**, 1157–1163.
- Donath, T. M., Natke, U., and Kalveram, K. T. (2002). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," *J. Acoust. Soc. Am.* **111**, 357–366.
- Eady, S. J. (1982). "Differences in the F_0 patterns of speech: Tone language versus stress language," *Lang. Speech* **25**, 29–42.
- Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions," *J. Acoust. Soc. Am.* **70**, 45–50.
- Francis, A. L., Clocca, V., Ma, L., and Fenn, K. (2008). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," *J. Phon.* **36**, 268–294.
- Gandour, J. T. (1983). "Tone perception in far Eastern languages," *J. Phon.* **11**, 149–175.
- Gandour, J. T., and Harshman, R. A. (1978). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," *Lang. Speech* **21**, 1–33.
- Guenther, F. H. (2006). "Cortical interactions underlying the production of speech sounds," *J. Commun. Disord.* **39**, 350–365.
- 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.
- Hawco, C. S., Jones, J. A., Ferretti, T. R., and Keough, D. (2009). "ERP correlates of online monitoring of auditory feedback during vocalization," *Psychophysiology* **46**, 1216–1215.
- Jones, J. A., and Keough, D. (2008). "Auditory-motor mapping for pitch control in singers and nonsingers," *Exp. Brain Res.* **190**, 279–287.
- Jones, J. A., and Munhall, K. G. (2002). "The role of auditory feedback during phonation: Studies of Mandarin tone production," *J. Phon.* **30**, 303–320.
- Kawahara, H. (1994). "Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies," *J. Acoust. Soc. Jpn.* **15**, 201–202.
- Lane, H., and Webster, J. W. (1991). "Speech deterioration in postlingually deafened adults," *J. Acoust. Soc. Am.* **89**, 859–866.
- Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (2008). "Interactions between auditory and somatosensory feedback for voice F_0 control," *Exp. Brain Res.* **187**, 613–621.
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (2001). "Comparisons of voice F_0 responses to pitch-shift onset and offset conditions," *J. Acoust. Soc. Am.* **110**, 2845–2848.
- Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (2000). "Effects of pitch-shift onset velocity on voice F_0 responses," *J. Acoust. Soc. Am.* **107**, 559–564.
- Larson, C. R., Sun, J., and Hain, T. C. (2007). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control," *J. Acoust. Soc. Am.* **121**, 2862–2872.
- Lee, Y. S., Vakoch, D. A., and Wurm, L. H. (1996). "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," *J. Psycholinguist. Res.* **25**, 527–542.
- Li, C. N., and Thompson, S. A. (1977). "The acquisition of tone in Mandarin-speaking children," *J. Chin. Linguist.* **4**, 185–200.
- Liu, H., Auger, J., and Larson, C. R. (2010). "Voice fundamental frequency modulates vocal response to pitch perturbations during English speech," *J. Acoust. Soc. Am.* **127**, EL1–EL5.
- Liu, H., and Larson, C. R. (2007). "Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex," *J. Acoust. Soc. Am.* **122**, 3671–3677.
- Liu, H., Xu, Y., and Larson, C. R. (2009). "Attenuation of vocal responses to pitch perturbations during Mandarin speech," *J. Acoust. Soc. Am.* **125**, 2299–2306.
- Natke, U., Donath, T. M., and Kalveram, K. T. (2003). "Control of voice fundamental frequency in speaking versus singing," *J. Acoust. Soc. Am.* **113**, 1587–1593.
- Natke, U., and Kalveram, K. T. (2001). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," *J. Speech Lang. Hear. Res.* **44**, 577–584.
- Oller, D., and Eilers, R. (1988). "The role of audition in infant babbling," *Child Dev.* **51**, 441–449.
- Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). "Plasticity in the adult human auditory brainstem following short-term linguistic training," *J. Cogn. Neurosci.* **20**, 1892–1902.
- Tse, J. K.-P. (1978). "Tone acquisition in Cantonese: A longitudinal case study," *J. Child Lang.* **5**, 191–204.
- Vance, T. J. (1976). "An experimental investigation of tone and intonation in Cantonese," *Phonetica* **33**, 368–392.
- Vance, T. J. (1977). "Tonal distinctions in Cantonese," *Phonetica* **34**, 93–107.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," *Nat. Neurosci.* **10**, 420–422.
- Xu, Y. (1997). "Contextual tonal variations in Mandarin," *J. Phon.* **25**, 61–83.
- 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.
- Zarate, J. M., and Zatorre, R. J. (2008). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," *Neuroimage* **40**, 1871–1887.