Brain & Language 121 (2012) 25–34

Contents lists available at SciVerse ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization

Zhaocong Chen[a,1], Peng Liu[a,1], Emily Q. Wang[b], Charles R. Larson[c], Dongfeng Huang[a], Hanjun Liu[a,*]

[a] Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, PR China
[b] Department of Communication Disorders and Sciences, and Otolaryngology, Rush University Medical Center, 1611 West Harrison St., Chicago, IL 60612, USA
[c] Department of Communication Science and Disorders, Northwestern University, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Accepted 3 February 2012
Available online 28 February 2012

Keywords:
Auditory feedback
Vocal pitch
Event-related potentials
Language experience
Auditory–vocal integration

ABSTRACT

The present study investigated whether the neural correlates for auditory feedback control of vocal pitch can be shaped by tone language experience. Event-related potentials (P2/N1) were recorded from adult native speakers of Mandarin and Cantonese who heard their voice auditory feedback shifted in pitch by −50, −100, −200, or −500 cents when they sustained the vowel sound /u/. Cantonese speakers produced larger P2 amplitudes to −200 or −500 cents stimuli than Mandarin speakers, but this language effect failed to reach significance in the case of −50 or −100 cents. Moreover, Mandarin speakers produced shorter N1 latencies over the left hemisphere than the right hemisphere, whereas Cantonese speakers did not. These findings demonstrate that neural processing of auditory pitch feedback in vocal motor control is subject to language-dependent neural plasticity, suggesting that cortical mechanisms of auditory–vocal integration can be shaped by tone language experience.

1. Introduction

Normal human vocal communication requires auditory feedback so that the speaker can insure that the production accurately reflects the intended sound, and if a mismatch is found between the intended and actual feedback, the vocal output can be corrected. One of the most important acoustical features of the voice, fundamental frequency (F_0), or its perceptual attribute pitch, is crucial for normal speech communication. Understanding mechanisms underlying the integration of auditory pitch feedback for the control of vocal production has been a long-standing goal for speech research.

Since 1981, the altered auditory feedback paradigm has been used to examine the mechanisms of pitch processing in auditory–vocal integration (Elman, 1981). In this paradigm, subjects hear their voice pitch feedback unexpectedly shifted upwards or downwards during vowel phonation or speech production. A general finding is that, when voice pitch feedback is perturbed, people adjust their vocalization by changing their voice pitch in the direction opposite to the stimulus (Burnett, Freedland, Larson, & Hain, 1998; Jones & Munhall, 2002; Larson, Sun, & Hain, 2007; Natke, Donath, & Kalveram, 2003). It has been suggested that auditory pitch feedback can be used to correct for pitch errors between the intended and actual feedback for stabilizing voice F_0 level at a desired level across different tasks (Chen, Liu, Xu, & Larson, 2007; Liu & Larson, 2007; Liu, Xu, & Larson, 2009) or subject populations (Kiran & Larson, 2001; Liu, Chen, Jones, Huang, & Liu, 2011; Russo, Larson, & Kraus, 2008). Recently, several electrophysiological and brain imaging studies have been conducted to examine how auditory pitch feedback is encoded at the cortical level during vocal pitch regulation. For example, event-related potential (ERP) studies have shown that greater N1/P2 amplitudes are associated with larger magnitudes of pitch feedback perturbation (Behroozmand, Karvelis, Liu, & Larson, 2009; Hawco, Jones, Ferretti, & Keough, 2009; Liu, Meshman, Behroozmand, & Larson, 2011). Using functional magnetic resonance imaging (fMRI), Zatorre and his colleagues (Zarate, Wood, & Zatorre, 2010; Zarate & Zatorre, 2008) reported that the anterior cingulate cortex, auditory cortex and putamen were activated to correct pitch error in auditory feedback during self-vocalization.

Unlike other acoustic cues such as duration or intensity, one of the most important attributes of pitch is to serve as a cue for tone identification. In tone languages such as Mandarin, variations in pitch are crucially important for identifying the meanings of otherwise identical syllables, while in non-tone languages such as English, they are used to signal stress and intonation patterns. Given the distinct role of pitch in different languages, mechanisms underlying the integration of auditory pitch information into the control of vocal production may be shaped by language experience. Numerous studies of pitch perception have demonstrated that pitch processing at the cortical/subcortical level is influenced by language experience (Bidelman, Gandour, & Krishnan, 2011; Chandrasekaran, Krishnan, & Gandour, 2007b; Krishnan & Gandour, 2009; Krishnan,

* Corresponding author. Fax: +86 20 87750632.
E-mail address: lhanjun@mail.sysu.edu.cn (H. Liu).
[1] These authors contributed equally to this study.

0093-934X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2012.02.004

Xu, Gandour, & Cariani, 2004, 2005; Xu et al., 2006). For example, relative to English subjects, Mandarin subjects showed larger mismatch negativity (MMN) responses in processing linguistically relevant pitch contours (Chandrasekaran, Krishnan, & Gandour, 2007a, 2007b, 2009). As reflected by the human frequency following response (FFR), previous studies comparing Mandarin with English show that native experience with lexical tones enhances pitch encoding at the brainstem irrespective of linguistic or non-linguistic context (Krishnan, Swaminathan, & Gandour, 2009; Krishnan et al., 2005; Swaminathan, Krishnan, & Gandour, 2008).

Evidence has shown that perception of auditory signals can affect speech production (Amir, Amir, & Kishon-Rabin, 2003; Kappes, Baumgaertner, Peschke, & Ziegler, 2009; Levelt, 1983; Watts, Moore, & McCaghren, 2005); in other words, speech perception is closely integrated with speech production (Hickok, Houde, & Rong, 2011; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Wilson, Saygin, Sereno, & Iacoboni, 2004). For example, non-musicians with better pitch discrimination sang more accurately than those with poorer pitch discrimination (Amir et al., 2003; Watts et al., 2005). Accurate voice F_0 production requires an integration between auditory perception and vocal production, during which unexpected pitch changes must be detected so that the mismatch between the intended and actual auditory feedback can be translated to the vocal motor systems for vocal output correction. Thus, it is reasonable to hypothesize that the saliency in perceiving small pitch changes may result in an enhanced ability for processing pitch errors in auditory feedback during the control of vocal production. Given the language-dependent plasticity in pitch encoding (Krishnan & Gandour, 2009), the language-experience-induced enhancement of neural processing of pitch perception may facilitate pitch encoding mechanisms underlying voice F_0 production.

Some recent behavioral data provide supportive evidence for this hypothesis. One cross-language study showed that Mandarin speakers produced significantly larger vocal responses than Cantonese speakers to large pitch perturbations, suggesting an effect of language experience on the behavioral control of auditory pitch feedback during self-vocalization (Liu et al., 2010). Other behavioral studies showed that, as compared to English speakers (Burnett et al., 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001), Mandarin/Cantonese speakers seemed to respond more quickly to pitch feedback perturbations (Liu et al., 2010). Considering the inherent difference between behavioral and neurophysiological responses, however, whether cortical processing of auditory pitch feedback is language-dependent for auditory–vocal integration is still an open question.

Therefore, the purpose of the present study was to examine whether neural processing of auditory pitch feedback during self-vocalization is shaped by language experience. Two of the most widely used Chinese languages, Cantonese and Mandarin, were chosen for the present study. To the best of our knowledge, no study has yet been conducted to explore the cortical pitch processing of auditory–vocal integration between two different tone languages. Cantonese has six contrastive lexical tones including high level, mid level, low level, high rising, low rising, and low falling, while Mandarin has four tones of high level, high rising, falling rising, and high falling. Previous behavioral studies have shown that Cantonese listeners are better at discriminating the average F_0 level (high or low) than Mandarin listeners (Tse, 1978), and Cantonese listeners are also better at discriminating Mandarin tones than Mandarin listeners are at discriminating Cantonese tones (Lee, Vakoch, & Wurm, 1996). These results suggest that Cantonese tones are more difficult to perceive than Mandarin tones, and Cantonese speakers may require more highly tuned perceptual abilities for tone discrimination than Mandarin speakers. Thus, we would expect to see that tonal language experience (Cantonese vs. Mandarin) can shape the neural processing of auditory pitch feedback during self-vocalization.

2. Methods

2.1. Subjects

Fifteen native Cantonese-speaking students (8 females and 7 males) and fifteen native Mandarin-speaking students (8 females and 7 males) from Sun Yat-sen University of China participated in the experiment. They were matched in age (Cantonese: 18–26 yr, mean = 21.5 yr, SD = 3.0 yr; Mandarin: 20–25 yr, mean = 21.4 yr, SD = 1.2 yr), and no significant difference was found in age between these two groups ($t = 0.139$, $p = 0.892$). Cantonese speakers spoke Guangzhou Cantonese and Mandarin, but they used Cantonese as their primary language in their daily life. All of the Mandarin speakers spoke Beijing Mandarin only and had no familiarity with Cantonese. All subjects were right-handed, and none of the subjects reported a history of speech, hearing, or neurological disorders. All subjects passed a hearing screening at the threshold of 25 dB HL for pure-tone frequencies of 0.5–4 kHz. They received monetary compensation for their participation and gave informed consent in compliance with a protocol approved by the Institution Review Board of The First Affiliated Hospital at Sun Yat-sen University of China.

2.2. Apparatus

All subjects were tested in a sound-treated booth throughout the experiment. Their voices were recorded by a Genuine Shupu microphone (model SM-306) and amplified by a MOTU Ultralite Mk3 firewire audio interface. Prior to the testing, acoustic calibration of the experimental system was performed to insure that the intensity of voice feedback heard by the subjects was 10 dB (SPL) higher than that of subject's voice output. An Eventide Eclipse Harmonizer controlled by a custom-developed MIDI software program (Max/MSP, v.5.0 by Cycling 74) was used to pitch-shift the amplified voices randomly. The pitch-shifted voices were played back to subjects through insert earphones (ER1, Etymotic Research Inc.). Transistor–transistor logical (TTL) control pulses signaled the onset and offset of the pitch-shift stimuli. The voice, feedback, and TTL control pulses were digitized at 10 kHz by a PowerLab A/D converter (model ML880, AD Instruments), and recorded using LabChart software (v.7.0 by AD Instruments).

2.3. Procedure

During the experiment, subjects were asked to vocalize a vowel sound /u/ for about 5–6 s at their habitual and comfortable pitch level. They were instructed to vocalize the vowel sound without contracting speech muscles by holding their face, jaw and tongue in a stationary position, which would minimize the effect of muscle contraction on the quality of the electroencephalogram (EEG) signals. During each vocalization, their voice feedback was randomly pitch-shifted downwards five times (see Fig. 1A), and production of 20 consecutive vocalizations constituted a block, resulting in a total of 100 trials per block. During each vocalization, the first pitch-shifted stimulus was presented with a delay of 500–1000 ms after vocal onset, and the succeeding stimuli occurred with an inter-stimulus interval (ISI) of 700–900 ms. For each block, the duration of the pitch perturbations was fixed at 200 ms and the magnitude was held constant at −50, −100, −200, or −500 cents. The reason for using downward pitch-shift stimulus was that, relative to upward direction, downward direction yields more prominent neural responses to pitch feedback perturbations (Liu et al., 2011). The