

Neural control of fundamental frequency rise and fall in Mandarin tones

Peter Howell^{a,*}, Jing Jiang^b, Danling Peng^b, Chunming Lu^{b,*}

^aDivision of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, England, UK

^bState Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, PR China

ARTICLE INFO

Article history:

Accepted 19 January 2012

Available online 15 February 2012

Keywords:

Mandarin tone

Tone rise

Tone fall

Neural control

Connectivity

ABSTRACT

The neural mechanisms used in tone rises and falls in Mandarin were investigated. Nine participants were scanned while they named one-character pictures that required rising or falling tone responses in Mandarin: the left insula and right putamen showed stronger activation between rising and falling tones; the left brainstem showed weaker activation between rising and falling tones. Connectivity analysis showed that the significant projection from the laryngeal motor cortex to the brainstem which was present in rising tones was absent in falling tones. Additionally, there was a significant difference between the connection from the insula to the laryngeal motor cortex which was negative in rising tones but positive in falling tones. These results suggest that the significant projection from the laryngeal motor cortex to the brainstem used in rising tones was not active in falling tones. The connection from the left insula to the laryngeal motor cortex that differs between rising and falling tones may control whether the rise mechanism is active or not.

Crown Copyright © 2012 Published by Elsevier Inc. All rights reserved.

1. Introduction

The larynx is involved in many aspects of speech production, including suprasegmental control, the transition from voiced to voiceless excitation and direction of fundamental frequency movement that is contrastive on segments as short as a syllable in tonal languages (Ludlow, 2005; Yip, 2002). Relatively little is known about neural control of the larynx in general and, of the functions indicated earlier, least is known about neural and motor control in the production of segmentally contrasting tones (Ludlow, 2005; Simonyan & Horwitz, 2011). This is surprising, as semantic contrasts based on tone differences are used in some languages, including Mandarin, the language spoken by the largest number of people in the world.

Four tones occur in Mandarin stressed syllables (designated T1, T2, T3 and T4), which arise mainly from the voice fundamental frequency movements that occur in the syllables. Idealized movements in fundamental frequency for the four tones are represented schematically in Fig. 1a. These movements alone are sufficient to give contrastive meanings between pairs of syllables (Lin, 2001).

Children learn to make tone contrasts at a later age than place, manner and voicing contrasts. T1–T4 fall into two groups based on the age at which they are acquired: T1 and T4 are acquired earlier than T2 and T3 (Hua & Dodd, 2000; Li & Thompson, 1977). This

suggests that members of the latter pair are difficult to produce in comparison to the former pair. Another fact pointing to T2 and T3 being more difficult than T1 and T4 is that most tone confusion errors involve T2 and T3 (Clumbeck, 1977; Li & Thompson, 1977). This may indicate that the ability to produce rises and falls is reliant on neural mechanisms that mature at different ages. This hypothesis is backed up by observations that show that different laryngeal maneuvers are associated with production of rising and falling voice fundamental frequency (Harvey & Howell, 1980; Ludlow, 2005).

The current study was designed to investigate whether different neural mechanisms give rise to fundamental frequency rises and falls, so T2 and T4 were the main focus. The evidence for the different laryngeal maneuvers that achieve fundamental frequency rises and falls is summarized next, followed by a review of the evidence for the neural mechanisms that control the larynx.

1.1. Laryngeal mechanisms responsible for fundamental frequency rises and falls

One theory is that voice fundamental frequency drops when activity ceases in the laryngeal structures responsible for achieving rises (Harris, 1974). If this was true, a single neural mechanism could be responsible for rise and fall maneuvers. However, one problem for this theory is why falls in fundamental frequency occur at different rates on different tones (as shown when T3 and T4 are compared in Fig. 1a). The majority of researchers consider that there are active fundamental frequency lowering mechanisms. These either use different muscles from those used for

* Corresponding authors. Fax: +44 (0) 20 7436 4276 (P. Howell), +86 (0) 10 5880 6154 (C. Lu).

E-mail addresses: p.howell@ucl.ac.uk (P. Howell), luchunming@bnu.edu.cn (C. Lu).

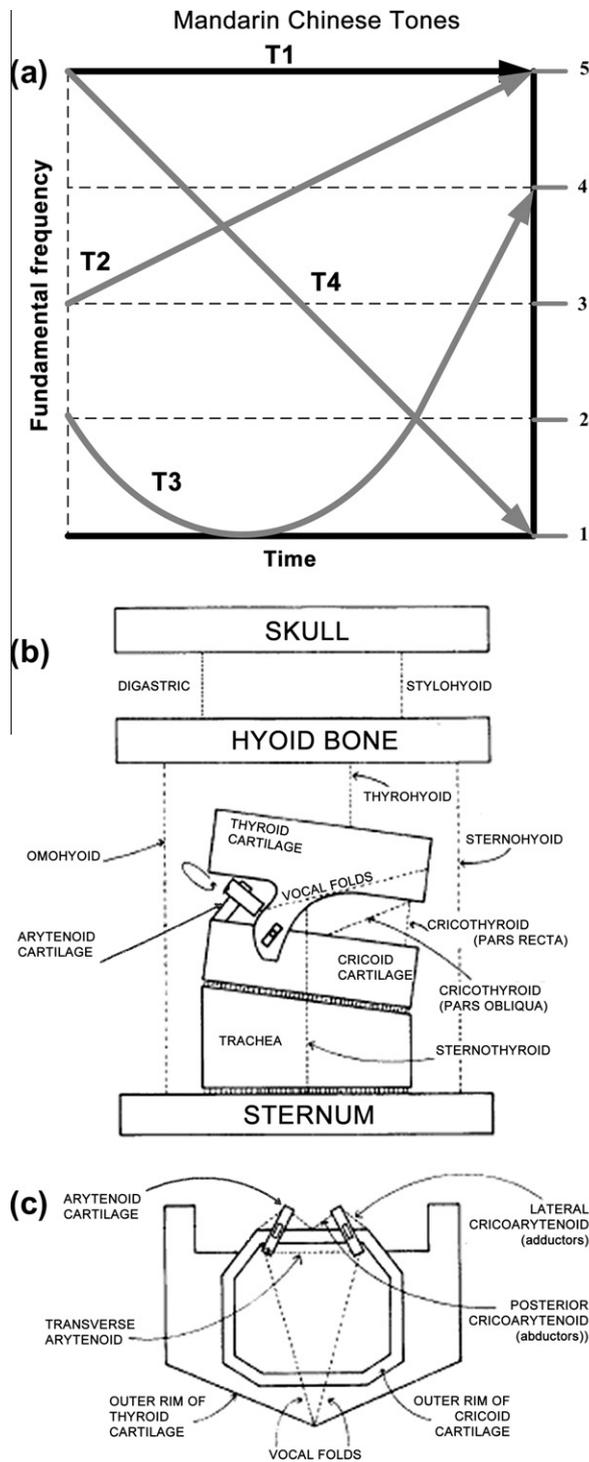


Fig. 1. Mandarin tonal movement. (a) The range of fundamental frequencies that speakers of Mandarin use can be split into five equally spaced sub-ranges. T1 starts and ends at level 5; T2 starts at level 3 and increases to level 5; T3 starts at level 2, dips to level 1, and then rises to level 4; T4 starts at level five and decreases to level 1. (b) and (c) the cartilaginous and muscular tissues that are important for achieving changes in voice fundamental frequency (reprinted from Harvey & Howell, 1980, with permission).

raising fundamental frequency, or employ the same muscles used when making rises but control them in different ways (Hirano, Vennard, & Ohala, 1970; MacNeilage, 1972; Ohala, 1970).

The larynx has nine cartilages altogether, three of which are especially important for fundamental frequency control (shown in Fig. 1b and c). Two of these are single cartilages (the thyroid

and cricoid) while the third is a bilateral pair of cartilages (the arytenoids). The thyroid cartilage sits above, and is larger in diameter than, the cricoid cartilage. The thyroid and cricoid cartilages can be moved up or down by extrinsic muscles, which are connected to the hyoid bone, skull and sternum. There are also several muscles that only connect with structures within the larynx (the intrinsic muscles).

Vocal fold tension determines fundamental frequency. The vocal folds run from the front of the thyroid cartilage to the vocal processes of the arytenoid cartilages, which in turn are seated on the rear of the cricoid cartilage. The vocalis muscle and other tissue form the body of the vocal folds. In voiced speech the vocal folds go through a cycle where air from the lungs forces them apart and then they come together again because of the drop in pressure, suction due to the Bernoulli eddying action of the released air and tissue elasticity. The vibration rate of the vocal folds determines the voice fundamental frequency, so it is necessary to understand how the cartilages and the intrinsic and extrinsic muscles alter their tension.

The cricothyroid is the main intrinsic laryngeal mechanism that raises and lowers fundamental frequency. As it contracts, it tilts the cricoid and thyroid cartilages, stretches the vocal folds and changes their frequency of vibration. The vibration rate of the vocal folds goes up when the cricothyroid is tensed and, conversely, the vibration rate goes down when the cricothyroid is relaxed. Support for this mechanism from tone languages is that the activity level in this muscle precedes rises in fundamental frequency (Yip, 2002, p. 8). Another intrinsic mechanism that produces fundamental frequency falls is contraction of the thyroarytenoid (Ohala, 1978).

The extrinsic laryngeal musculature may also be used to change fundamental frequency by changing the position of the thyroid cartilage, which changes the length and anterior–posterior tension (Zenker, 1964) or vertical tension (Ohala, 1972) of the vocal folds. Of the extrinsic muscles, the suprahyoid muscles increase fundamental frequency whereas the infrahyoid muscles decrease voice fundamental frequency. Erickson (1993) found extrinsic muscle activity corresponding to the initial fall in the rising tone in Thai.

It should be cautioned that voice fundamental control is more complex than described here. For instance, the cricothyroid and thyroarytenoid muscles act synergistically to achieve changes in fundamental frequency (Ohala, 1978; Titze, Luschei, & Hirano, 1989) and the role of the cricothyroid depends upon vocal fold position at the time of contraction (Kuna, Smickley, Vanoye, & McMillan, 1994; Titze et al., 1989). Fundamental frequency rises and falls can also be achieved by increases or decreases of subglottal air pressure (Herman, Beckman, & Honda, 1996; Monsen, Engbretson, & Vemula, 1978). Subglottal pressure changes are mainly achieved by muscles that adjust the pulmonary system. Nevertheless, the earlier statements about how the structures in and around the larynx affect voice fundamental frequency are the main influences and show that different muscles are responsible for voice fundamental rises and falls. In turn, these observations suggest that different neural mechanisms may be involved when T2 and T4 are produced, which corresponds with the conclusion based on the difference in age of acquisition of these two tones. Work on possible neural mechanisms that control laryngeal activity is reviewed next.

1.2. Neural mechanisms for controlling voluntary laryngeal activity

Animal and imaging evidence have shown that two parallel pathways are implicated in voluntary laryngeal control: (1) the anterior cingulate cortex (ACC)-periaqueductal gray (PAG)-brainstem pathway, which controls the initiation of basic vocal reactions; and (2) the laryngeal-motor cortical pathway, which controls voluntary voice production (Simonyan & Horwitz, 2011). These

two pathways converge in the ACC and the brainstem, and together they allow appropriate coordination of learned vocal patterning and voice initiation (Hannig & Jürgens, 2006; Simonyan & Jürgens, 2002, 2005).

The laryngeal motor cortex (LMC), which is important for making fundamental frequency rises and falls, is located in the ventral part of the premotor cortex (BA4) (Brown, Ngan, & Liotti, 2008; Simonyan & Horwitz, 2011). Patients with damage to the LMC are occasionally able to initiate phonation that results in grunts, wails, and laughs. However, they cannot make voluntary modulations of pitch, intensity, and voice quality (Jürgens, 2002). Recent neuroimaging evidence has helped to further delineate the function of the LMC. For example, it has been shown that the LMC is selectively involved when syllable sequencing and syllable complexity differ (Bohland & Guenther, 2006). Motor control of laryngeal muscles when intonation varied induced stronger and more reliable activation in the LMC than motor control of laryngeal muscles without intonation (Olthoff, Baudewig, Kruse, & Dechent, 2008).

Recent investigations have shown that humans and nonhuman primates share a common network of extensive cortical and sub-cortical connections with the LMC (Simonyan & Horwitz, 2011). Most of the connections with the LMC are bi-directional, including connections with the insula, the surrounding somatosensory cortex, inferior frontal cortex, cingulate cortex, and inferior parietal cortex including the angular gyrus (AG) and the supramarginal gyrus (SMG). These regions are involved in the integration of proprioceptive and tactile feedback, monitoring of verbal responses and motor preparation and processing (Fiebach, Friederici, Smith, & Swinney, 2007; Peschke, Ziegler, Kappes, & Baumgaertner, 2009; Simonyan, Ostuni, Ludlow, & Horwitz, 2009). A few of the connections are uni-directional, such as the projections from the LMC to the putamen, the caudate nucleus, and the brainstem nuclei. The uni-directional connections are associated with integrative control of different aspects of speech production, ranging from motor control to motivation and cognitive processing of speech (Jürgens, 2002; Jürgens & Ehrenreich, 2007). Neuroimaging evidence indicates that the LMC network shows significant left-hemispheric lateralization during voice production but not during controlled breathing (Simonyan et al., 2009).

1.3. Functional considerations

To date, there is no specific imaging evidence of how the neural systems responsible for control of the larynx are related to control of fundamental frequency rises and falls in tonally-contrasting material. There are studies that suggest what regions might be implicated in functional control of both rises and falls based on task analyses, and some non-imaging studies that suggest which regions may be specifically associated with rise or fall control. Each of these is dealt with in turn.

At a general level, the fundamental frequency rises and falls require sequential control of muscle activity. The projections from the LMC to the basal ganglia and the brainstem that achieve sequence-control may be active in different ways on rises and falls, as these maneuvers require different sequences of muscular adjustment. T2 and T4 are acquired at different ages, which suggest that they have different levels of phonological complexity. Consequently, the insula and inferior parietal cortex (AG and SMG) may be involved (Kast, Bezzola, Jancke, & Meyer, 2011; Zheng, Munhall, & Johnsrude, 2010).

Although few studies have specifically examined fundamental frequency rise and fall, other work has shed some light on the neural control of these maneuvers. For example, neuroimaging studies on singing indicated that, compared with voluntary vocal pitch regulation, involuntary vocal pitch regulation elicited higher

activity in several brain regions including the bilateral BA 6/44 and anterior insula (Zarate, Wood, & Zatorre, 2010). The brainstem is involved in the perceptual processing of tone (Krishnan, Gandour, & Bidelman, 2010). Moreover, compared with comfortable pitch-level production, high pitch-level production induced higher activation in the bilateral cerebellum, left inferior frontal gyrus, and left cingulate gyrus (Peck et al., 2009). Similar comparisons for low pitch-level production showed higher activation in the inferior frontal gyrus, insula, putamen, and cingulate gyrus in the left hemisphere (Peck et al., 2009). Since breathing control may be also involved in fundamental frequency rise and fall, the inferolateral sensorimotor cortex, premotor cortex, supplementary motor area, and striatum, which are involved in volitional inspiration, would also be expected (Evans, Shea, & Saykin, 1999).

The clinical literature was explored to see whether there was evidence for differential involvement of particular regions of the brain for T2 and T4 production as opposed to T1 and T3. Two topics that were examined, for which there was no literature concerning the effects on tone production, were drug studies and genetic disorders. Some production studies were found that indicated how fundamental frequency was affected by lesions. Many studies have shown that left hemispheric lesion will lead to deficit of both tone production and perception, whereas lesions to the right hemisphere do not (Gandour et al., 1992; Packard, 1986). It has been shown that left hemisphere brain damage affects all tone categories (Gandour & Dardarananda, 1983; Gandour, Petty, & Dardarananda, 1988), probably because lesions usually affect large regions of the brain. Focal lesions to the left parietal lobe affected the production of T2 and T3 more than that of T4, causing T2 and T3 to sound more like T1 (Wang, 2004). These lesion data are consistent with the fact that age of acquisition is similar for T2 and T3. Broca's and conduction aphasia affect production of all tones, but T3 is affected the most (Shi & Li, 2011). To summarize, the balance of evidence suggests that T3, and to a lesser degree T2, can be specifically affected by lesions, much more so than T1 and T4.

1.4. Summary and hypotheses

Whilst there are documented differences between the laryngeal maneuvers responsible for raising and lowering of voice fundamental frequency, the neural mechanisms responsible for achieving these maneuvers have not been documented. Neural circuitry that is likely to be involved in general aspects of laryngeal control, based mainly on animal studies, was identified by Simonyan and Horwitz (2011). We also identified most of the same regions as Simonyan and Horwitz in the functional task analysis at the start of the previous section. Together, these lines of evidence suggest which regions may be involved in voluntary laryngeal activity control. These indicate that the regions of interest are the LMC and brain regions that connect with it, such as the inferior frontal cortex, insula, inferior parietal cortex, putamen, and brainstem (Simonyan & Horwitz, 2011).

The current study examined central nervous system activity in Mandarin speakers whilst they produced syllables with rising or falling tones (T2 and T4 respectively). The purpose was to identify possible differences in the neural control of rises and falls. The two other tones were used in validations. T1 involves neither rise nor fall and T3 involves both a rise and fall, although in the case of T3, the age of acquisition and neurological evidence suggest that T3 is close to T2. The steps in the analysis were to examine the whole brain when rises and falls were produced, then to look at specific regions associated with rises and falls and, finally, structural equation modeling analysis was conducted to identify any different connectivity patterns for rises (T2) and falls (T4). Based on Simonyan and Horwitz (2011), it was expected that there would be bi-directional connections between the LMC and inferior frontal

cortex/insula and inferior parietal cortex, and uni-directional projections from the LMC to the putamen and brainstem. These connections may be responsible for achieving fundamental frequency rises and falls.

2. Materials and methods

2.1. Participants

Nine participants (five males and four females) were recruited from Beijing Normal University. They reported that they had no history of language, motor, or other neurological diseases. Their mean age was 24 years (the range being from 22 to 29 years). The mean number of years they had been in education was 15.5 (the range being from 12 to 19 years). All participants were right-handed, native Mandarin speakers, where handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971); A cutoff score of +40 was used as an indication of right handedness. The study was approved by the ethics committee of the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. Written informed consent was obtained from each participant before the experiment.

2.2. Experimental tasks and materials

Participants were scanned while they performed a picture-naming task. Forty-eight simple line drawings of common objects were selected from a standardized picture database (Zhang & Yang, 2003). The objects in each picture had a common Mandarin name that was one character long. The name of the object had one of the four tones (T1–T4) and there were 12 pictures for each tone. Forty-eight control images were formed to provide a baseline condition. These were generated by randomizing the pixels of each of the original 48 pictures. These control images were not namable. The 48 namable pictures and 48 unnamable control pictures were randomly presented to the participants in an event-related design in one scanning run (see Fig. 2).

On each trial, a picture was presented for 1 s, and then a blank screen appeared that lasted for 2 s. When a picture was presented on the screen, the participants were asked to name it aloud as fast and accurately as possible. During the baseline (control) trials, participants were asked simply to view the unnamable control pictures and not to make any mouth movements.

A Pentium III-based notebook with the Inquisit software package (Inquisit 2.0.4.1230, 2004, Seattle, WA: Millisecond Software) controlled stimulus presentation. An LCD projector running in 1024 × 768 mode displayed stimuli from inside the MR control room onto a back-projection screen located at the foot of the MR

scanner. Participants viewed the stimuli via a mirror attached to the head coil above their eyes. Participants were familiarized with the stimulus presentation and response collection setup before they commenced the experiment. At the beginning of the familiarization process, the participants made fewer than 10% errors on each tone category, whereas after familiarization, they performed the task with no errors.

Age of acquisition (AoA) of the object-names was measured on a six-point scale based on a Mandarin modification of the standard Cortese and Khanna (2008) scale. The scale was changed from seven-point to six-point and the corresponding AoAs were 0–4, 4+–6, 6+–8, 8+–10, 10+–12 and 12+ years. The differences in AoA across the four tones were not significant, $F(3,44) = 1.566$, $p = 0.211$. Therefore the words would have been acquired by participants at a similar age. Word frequency was obtained from the Modern Chinese Information Dictionary (Institute of Linguistics, 1986). Word frequency did not differ significantly across the four sets of tones, $F(3,44) = 1.323$, $p = 0.279$. Familiarity was obtained from a standardized picture database (Zhang & Yang, 2003). Word familiarity did not differ significantly across the four types of tone $F(3,44) = 0.299$, $p = 0.826$. Mean and standard deviation for AoA, word frequency and familiarity for the four tone types, T1–T4, are summarized in Table 1.

2.3. Imaging data acquisition

Imaging data were acquired with a 1.5 T whole-body Siemens Magnetom Sonata Maestro Class scanner (Siemens, Erlangen, Germany) equipped with the standard clinical head coil. During the experiment, participants lay supine within the MR scanner with their head secured by foam padding for the entire experimental run. MRI compatible headphones were worn to reduce the background noise.

Functional whole-brain T2-weighted images were acquired using a single-shot gradient-recalled echo-planar imaging (EPI) sequence. The parameters were time repetition, TR = 3000 ms; time echo, TE = 50 ms; flip angle = 90°; field of view, FOV = 220 mm, matrix = 64 × 64 (in-plane resolution = 3.4 × 3.4 mm), 20 slices, slice thickness = 6 mm and slice acquisition was interleaved.

Table 1
Summary of AoA, word frequency and familiarity of each word set.

	T1	T2	T3	T4
AoA	2.442(0.170)	2.783(0.198)	2.638(0.198)	3.055(0.251)
Frequency	0.050(0.013)	0.120(0.065)	0.316(0.166)	0.176(0.078)
Familiarity	4.548(0.101)	4.383(0.169)	4.528(0.181)	4.413(0.136)

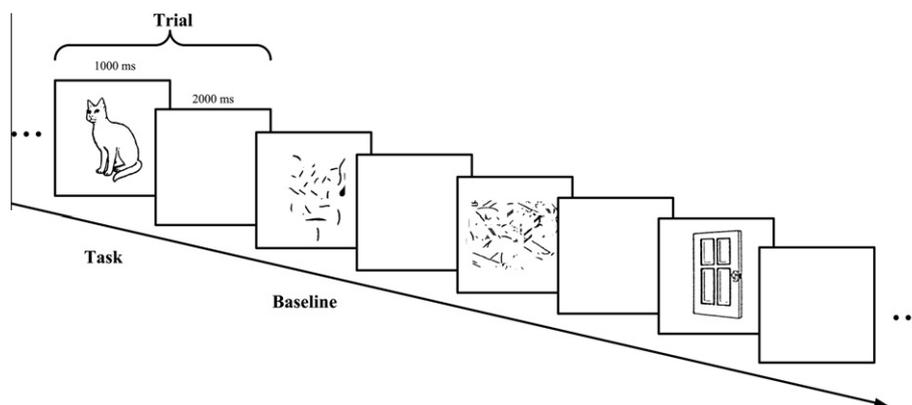


Fig. 2. Experimental protocol.

For anatomical localization, standard whole-brain, high-resolution 3D structural images were acquired after the functional scan using a T1-weighted MP-RAGE sequence (TR = 1970 ms; TE = 3.93 ms; flip angle = 15°; FOV = 220 mm; matrix = 256 × 256; 96 slices; slice thickness = 1.7 mm, sagittal plane; resolution = 0.48 × 0.48 mm).

2.4. Imaging data analysis

2.4.1. Preprocessing

The data were processed using Analysis of Functional Neuro-Images software AFNI, obtained from <http://afni.nimh.nih.gov/afni> (Cox, 1996). The first two volumes of the EPI images were discarded to allow for stabilization of the magnetic field. Six movement parameters were generated, which were used in the following individual level statistical analysis to exclude potential movement artifacts. Slice-time correction, image registration and motion correction were then performed using AFNI. The functional image time series were smoothed by low-pass filtering and application of an Isotropic Gaussian blur (full width at half maximum = 6 mm).

2.4.2. Individual level statistics

Statistical analysis of the individual functional imaging data was performed using generalized linear modeling (GLM) methods. Regression coefficients, β , were obtained for each tone category by deconvolving the measured time series using a Legendre polynomial fitting method. The coefficients were then converted into percent signal change. The percent signal change (converted β weights) provided an indication of the functional activation in response to the task for each participant, as compared with the baseline (task minus baseline). The six estimated motor parameters were used to exclude potential movement artifacts in the GLM model. Finally, individual images were normalized into MNI (Montreal Neurological Institute) space.

2.4.3. Group level differences in brain activations for T1–T4

For the group level statistical tests of brain activations, activity associated with each tone was computed first. These analyses used one-sample *t*-tests ($p < 0.05$, corrected by using Monte Carlo simulation, with individual voxel $p < 0.005$ and cluster size $> 218 \text{ mm}^3$) (Forman et al., 1995; Xiong, Gao, Lancaster, & Fox, 1995). Subsequently, paired-sample *t*-tests were conducted to establish differences between selected tone categories ($p < 0.05$, corrected by using Monte Carlo simulation, with individual voxel $p < 0.01$ and cluster size $> 321 \text{ mm}^3$).

2.4.4. Connectivity analysis using structural equation modeling (SEM)

2.4.4.1. Mode setup. The brain regions that showed activation differences between T2 and T4 (the tones of primary interest), i.e., the left insula and brainstem, and the right putamen, were examined in the SEM model. Additionally, since previous studies have shown that laryngeal activity is controlled by the LMC, this region was also selected. The LMC was localized based on previous literature (Brown et al., 2008; Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Simonyan & Horwitz, 2011). It should be noted that the LMC location also showed activation for T2 and T4 in this experiment (see Fig. S1). The connections between these brain regions were examined using SEM. Based on previous literature considered in the introduction, it was expected that there would be bi-directional connections between the LMC and insula, and uni-directional projections from the LMC to the putamen and brainstem (Simonyan & Horwitz, 2011).

2.4.4.2. Preprocessing for SEM. Within each brain region, the averaged time series across the voxels was extracted first in a sphere with a 3 mm radius, centered at the coordinates of the maximum

value within each region by using the AFNI program. Then, time points that corresponded to T1/T3/T4 were removed, leaving the time points that corresponded with T2. A similar procedure was applied to T4 (with T1/T2/T3 removed). Although this approach would tend to lose the temporal information in the data, this would not affect SEM results because SEM does not consider temporal information. Finally, principal components analysis was used to identify the “average” pattern of responses in each ROI across all participants in each tone category (Büchel, Coull, & Friston, 1999).

2.4.4.3. Model estimation. LISREL 8.7 (www.ssicentral.com) was used to estimate the parameters for the SEM model. An iterative maximum likelihood algorithm was used to calculate path coefficients and to achieve the best match between the covariance matrix reproduced by the model and the observed variance-covariance structure in the data (Jöreskog, 1996). The maximum likelihood (ML) discrepancy function was used to indicate the fit of the model, which yielded an overall fit statistic that follows a χ^2 -distribution under the null hypothesis assumption that the model correctly represents the data. In addition to the ML discrepancy function, other alternative fit indices were examined including the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Parsimony Goodness of Fit Index (PGFI) (Browne & Cudeck, 1993). These have been used in neuroimaging studies previously (Bentler, 1990; Honey et al., 2003). Besides the overall fit indices, the reported *t* value for each path coefficient in the model should be greater than a specified critical value to reject the null hypothesis that the path coefficient was 0. A path coefficient threshold of 0.05, corrected for false discovery rate (FDR) was used (Genovese, Lazar, & Nichols, 2002).

Statistical inferences about tone differences were based on a stacked-models approach. This started with a free model, in which all path connections were allowed to vary when the two tones T2 and T4 were presented. Then a restricted mode was developed in which a specified connection was constrained to be equal for the two tones. To do this, first, an omnibus test was applied in which the model with all parameters constrained to be the same for the two tones (constrained model) was compared with the model without any constraints (free model). This step showed whether any of the paths between the two models were significantly different for T2 and T4, but did not specify which of the paths were actually significantly different. At this stage, the comparison of models was done by subtracting the goodness-of-fit χ^2 value for the constrained model from the χ^2 value for the free model. The difference (χ^2_{diff}) was assessed with the degrees of freedom equal to the difference in the degrees of freedom for the constrained and free models (McIntosh et al., 1994). A significant χ^2_{diff} indicated that at least one path differed significantly across the two groups of tones (McIntosh & Gonzalez-Lima, 1994a,b).

When this omnibus test showed a significant difference between the two tones, the next step was to find which specific paths differed. This was done by constraining one path at a time to be the same between the two tones while other paths were unconstrained (estimated freely). This model was compared with the model without any constraints by a χ^2_{diff} ($df = 1$) test. A *p* value of 0.05 (two-tailed) was chosen as the threshold for significance (FDR corrected).

3. Results

Brain activations that were associated with T2 and T4 production were located first. Then, brain activation differences between T2 and T4 were computed. Regions that showed differences may indicate regions that specifically control rise and fall of fundamental frequency. Additional comparisons of T2 and T4 with T1 and T3

were used to validate the view that the activations were associated with rise and fall in the regions where there were differences in activity when T2 and T4 were compared. As first approximations, it was assumed that: (1) T1 (high-level tone) should not involve either the regions responsible for rises or those responsible for falls; (2) T3 (fall-rise tone) was similar to T2. The final analysis examined the connectivity patterns among brain regions that were involved in control of rise and fall of fundamental frequency.

3.1. Differences between T2 and T4

In general, similar neural activity patterns occurred during T2 and T4 articulation (see Fig. S2). However, some differences in activity patterns were also observed: The left insula (BA13, $x, y, z = -43, 12, -10, t = 3.625$, cluster size = 420 mm^3) showed stronger activation during T2 than during T4. A region located at the boundary between the left brainstem and cerebellum ($x, y, z = -16, -36, -43, t = -6.711$, cluster size = 445 mm^3) showed weaker activation during T2 than during T4 (see Fig. 3a). Further visual inspection was made to identify the location of this cluster in individual participants. It was found that among all nine participants, six participants' activations were in the brainstem, whereas the activations in the remaining three participants were in the cerebellum (these applied to both T2 and T4). As Fig. S3 shows, the brain region that was activated, involved both the cerebellar region that receives input from the frontal and parietal cortex and the brainstem and the brainstem region that has input to the cerebellar re-

gion. Another region in the right putamen ($x, y, z = 25, -1, -11, t = 3.828$, cluster size = 120 mm^3) also showed stronger activation during T2 than during T4 (see Fig. 3b) though it did not survive the threshold correction.

3.2. Differences between T2/T4 and T1

Based on the assumption made about T1 at the start of the results, subtracting T1 activity from T2 should not affect the pattern of neural activity responsible for rises (the left insula and the right putamen) and subtracting T1 activity from T4 should not affect the pattern of neural activity responsible for falls (the left brainstem).

Comparison between T2 and T1 showed that the left insula ($x, y, z = 43, 18, 10, t = 4.51$, cluster size = 140 mm^3) and the right insula ($x, y, z = 41, 20, 13, t = 4.212$, cluster size = 140 mm^3) showed stronger activations during T2 than during T1 ($p < 0.01$, uncorrected). No brain regions showed stronger activations during T1 than during T2 (see Fig. 4a). The T2–T1 comparisons results where T2 activity was higher than T1 support the association of left (and possibly right) insula activation with rising fundamental frequency. These comparisons also suggest that right insula activity is more labile (not observed when T2 and T4 were compared). The lack of any regions that showed stronger activity in T1 than T2 also supports the interpretation that brain regions responsible for controlling rises were identified.

Comparison between T4 and T1 showed that the left brainstem/cerebellum (brainstem, $x, y, z = -8, -42, -47, t = 5.671$, cluster

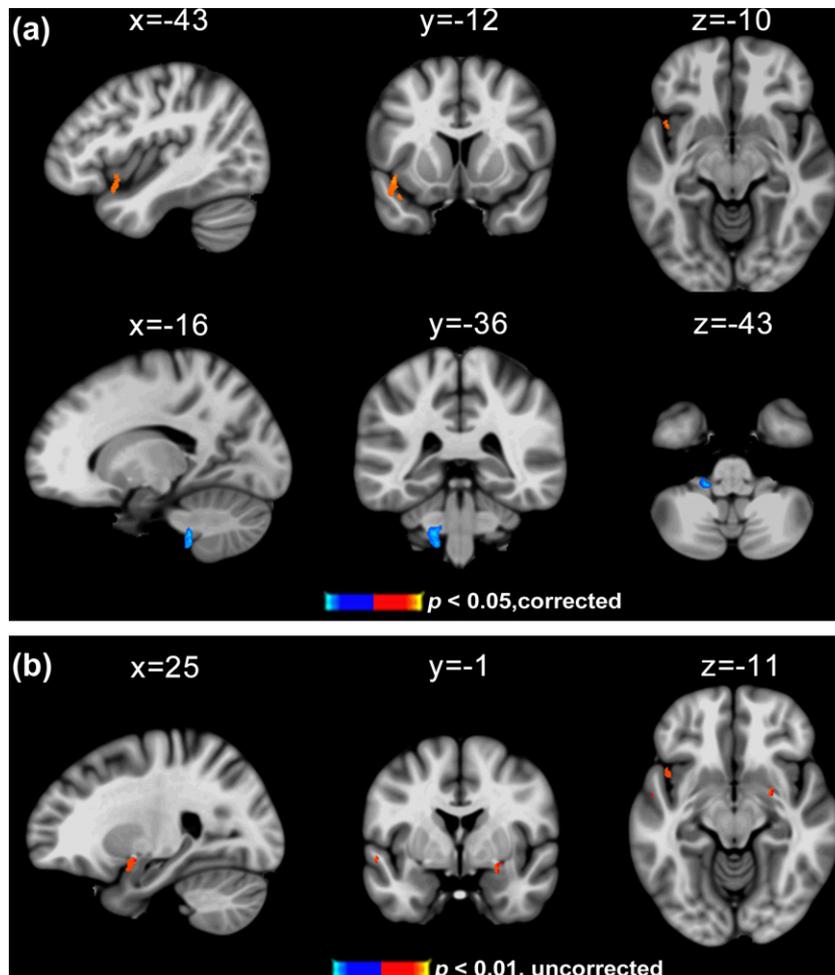


Fig. 3. Brain regions that showed activation differences between T2 and T4. (a) Above threshold brain regions ($p < 0.05$, corrected) and (b) below threshold regions ($p < 0.005$, uncorrected). Warm blobs indicate $T2 > T4$, cold blobs indicate $T2 < T4$.

size = 364 mm³) showed stronger activation in T4 than in T1. This location was very close to the region which showed weaker activation in T2 than T4 (see the comparison between T2 and T4 above). Additional stronger brain activation in T4 than in T1 was found in the middle temporal gyrus (BA21, $x, y, z = -59, -29, -8, t = 4.989$, cluster size = 803 mm³). No brain regions showed stronger activations in T1 than in T4 (see Fig. 4b). The T4–T1 comparison results support the association of left brainstem activation with falling fundamental frequency and suggest additional potential involvement of the left middle temporal gyrus. The lack of any regions that showed stronger activity in T1 than T4 also supports the interpretation that brain regions responsible for controlling falls were identified.

3.3. Differences between T3 and T2/T4

If T3 is similar in its neural processing to T2, as the age of acquisition and neurological evidence suggest, subtracting T3 neural activity from T2 should partially cancel activity in regions responsible for rises whereas subtracting T3 activity from T4 should not tend to affect activity associated with falls.

Comparison between T3 and T2 showed that only the right cerebellum (VIII) ($x, y, z = 23, 67, -56, t = -7.002$, cluster size = 560 mm³) showed weaker brain activation during T3 than during T2 (see Fig. 5a and Table 2). As all regional activity associated with T2 in the first analysis was obliterated and the only activity that arose was in a region not observed with either T2 or T4, it may be concluded that regional activity associated with rises had been appropriately located based on the T3 validation that assumes its activity pattern is similar to that of T2.

Comparison between T3 and T4 showed that the right middle frontal gyrus (BA46), right superior temporal gyrus (BA22, BA39), and left insula (BA13) showed stronger activations, whereas the right medial frontal gyrus (BA10) and left middle temporal gyrus (BA21) showed weaker activations during T3 than during T4 (see Fig. 5b and Table 2). Thus, T3 may be more similar to T2 because it shows a pattern similar to that observed when T2 and T4 were compared (the left insula that showed greater activity in T2 than T4, and no activity in the brainstem). The pattern of results would also be consistent with falls being due to a passive mechanism (no neural effects associated with falls), and/or falls being a result of a passive cessation of activity in the mechanisms that lead to rises.

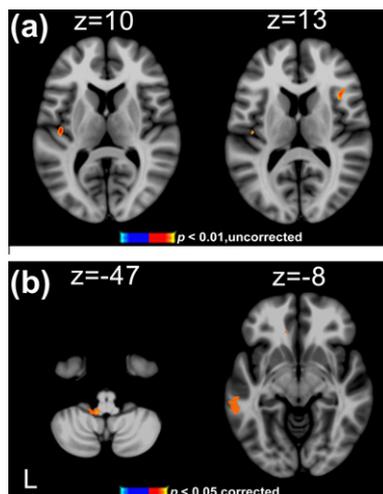


Fig. 4. Brain regions that showed activation differences between T2 and T1 (a) and between T4 and T1 (b). $p < 0.05$, corrected. Warm blobs indicate T2/T4 > T1, cold blobs indicate T2/T4 < T1.

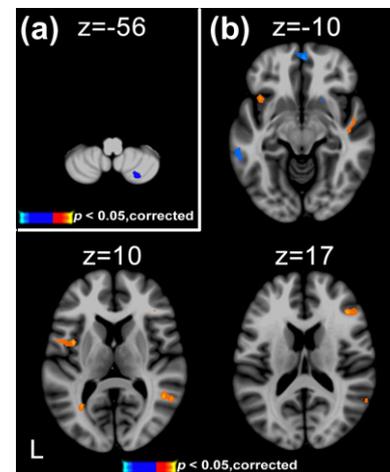


Fig. 5. Brain regions that showed activation differences between T3 and T2 (a) and between T3 and T4 (b). $p < 0.05$, corrected. Warm blobs indicate T3 > T2/T4, cold blobs indicate T3 < T2/T4.

The comparisons involving T2 and T4 with T1 and T3 provided some validation of the results when T2 and T4 alone were compared. The findings with the latter were taken as identification of regions responsible for controlling rises and falls respectively.

3.4. SEM results

The SEM procedure described in the method was used to investigate the connectivity patterns in the brain regions that were involved in the differences between T2 and T4.

3.4.1. Achievement of the best match between the model and the data

The SEM results showed that the *a priori* defined model was a good fit to the data of both T2 ($\chi^2 = 3.58, df = 6, p = 0.73$) and T4 ($\chi^2 = 0.51, df = 6, p = 1.00$). The overall statistical fit index based on the χ^2 value indicated that the covariance matrix was reproduced well by the model that matched best with the observed variance–covariance structure from the data (Jöreskog, 1996). This result was confirmed by other overall fit statistical indices (T2, RMSEA = 0.0, PGFI = 0.58, CFI = 1.00; T4, RMSEA = 0.0, PGFI = 0.60, CFI = 1.00). The standardized path coefficients for the best fitting model for each tone category (headed “T2” and “T4” in Table 3) and the overall fit indices (section headed with “comparison” in Table 3) are summarized in Table 3. The reported t value of these path coefficients showed that the connection from the insula to the LMC was significant for both T2 (negative connection) and T4 (positive connection), but the connection from the LMC to the insula was not. Additionally, the projection from the LMC to the brainstem was significant in T2, but not in T4. The projection from the LMC to the putamen was not significant for either T2 or T4.

3.4.2. Stacked model comparison

3.4.2.1. The omnibus test. As stated in the method section, the first step in the stacked model comparison was an omnibus test in which the model with all parameters constrained to be the same for the two tones (constrained model) was compared with the model without any constraints (free model). The results showed significant difference in path coefficients ($\chi^2_{diff} = 197.27, df = 4, p < 0.0001$) between T2 and T4, which indicated that at least one of the paths was significantly different between the two tones, but it was not known specifically which of the paths were actually significantly different. This was examined in the individual path test.

Table 2
Brain activation differences between T3 and T2/T4.

Brain area	Position			t-value	Cluster volume (mm ³)
	x	y	z		
<i>Tone3 > Tone2</i>					
Right Cerebellum (VIII)	–23	67	–56	–7.002	560
<i>Tone3 < Tone2</i>					
None					
<i>Tone3 > Tone4</i>					
Right middle frontal gyrus (BA46)	–43	–31	17	6.210	509
Right superior temporal gyrus (BA22)	–51	5	–8	4.449	401
Right superior temporal gyrus (BA39)	–47	51	13	9.505	882
Left insula (BA13)	40	–17	–11	5.601	540
<i>Tone3 < Tone4</i>					
Left middle temporal gyrus (BA21)	62	35	–10	–3.979	337

Note: The coordinates were standard MNI coordinates.

Table 3
Standardized path coefficients for T2 and T4, and results of individual path coefficients comparison between T2 and T4.

Paths	T2			T4			Comparison	
	Standard path coefficient	T	P	Standard path coefficient	t	P	χ^2_{diff}	P
LMC → Insula	0	0.00	1.000	0.00	0	1.000	0.00	1.000
Insula → LMC	–0.79	–7.21	0.000	0.7	5.11	0.000	81.04	0.000
LMC → Putamen	–0.26	–1.71	0.092	0.04	0.24	0.811	2.25	0.134
LMC → Brainstem	0.4	2.8	0.007	–0.15	–1	0.321	6.28	0.012

Note: The bold number indicated statistically significant path coefficients and significant differences between T2 and T4 (FDR corrected).

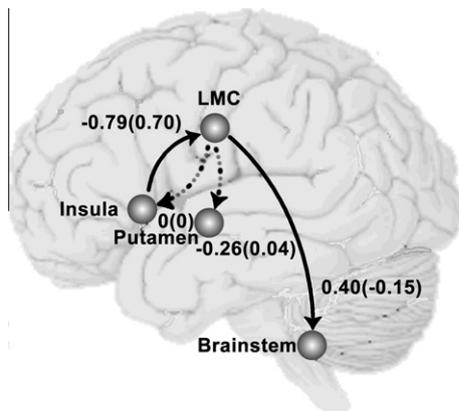


Fig. 6. Connectivity differences between T2 and T4. Solid and dash lines indicate significant and non-significant differences between T2 and T4, respectively. The number indicates standard path coefficients for T2 (outside the bracket) and T4 (inside the bracket). $p < 0.05$, corrected.

3.4.2.2. Individual path test. Specific paths that differed between T2 and T4 were located by examining the individual path coefficients (Fig. 6). This procedure revealed that the connection between the left insula and the left LMC and between the left LMC and the brainstem differed across the two tones. The connectivity from the LMC to the insula/putamen did not differ significantly between T2 and T4. Taken in conjunction with the earlier whole brain analyses, T2 appears to be predominantly controlled by the insula to LMC projection path, and T4 to be predominantly controlled by the LMC to brainstem path (see Table 3). Possible ways in which these pathways could function are examined in the discussion.

4. Discussion

The current study examined which neural mechanisms give rise to fundamental frequency rises and falls. One point of view is that a

single neural mechanism might be responsible for rise and fall maneuvers so that voice fundamental frequency drops when activity ceases in the laryngeal mechanism responsible for achieving rises (Harris, 1974). On the other hand, the majority of researchers consider that there are active fundamental-frequency lowering mechanisms. The present results support both of these views to some extent. Specifically, both rise and fall are controlled by the left LMC and brain regions that are connected with it. However, compared with T2, the projection of the LMC to the brainstem disappeared in T4, suggesting that the significant projection used in T2 may have ceased its activity in T4. An indication to cease activity that controls T2 may have resulted from an active change in the connection from the left insula to the LMC in T4 compared to that in T2. The functioning of the neural mechanisms associated with T2 and T4 are discussed next.

4.1. Neural mechanism for tone rise

4.1.1. Regionally stronger neural activation in T2 than in T4

Although a similar neural activation pattern was found in the production of both T2 and T4, the direct comparison between T2 and T4 revealed one brain region, the left anterior insula, that showed stronger neural activity during T2 production than during T4 production. Based on clinical and functional imaging data, the left anterior insula has been assumed to support prearticulatory functions of speech motor control such as the programming of vocal tract gestures, whereas other evidence suggests this region contributes to the actual coordination of the up to 100 muscles engaged in articulation and phonation (Ackermann & Riecker, 2004, 2010). Furthermore, the left anterior insula has been reported to be the only region with decreased cortical thickness in spasmodic dysphonia, which is a primary focal dystonia characterized by involuntary spasms in the laryngeal muscles during speech production (Simonyan & Ludlow, 2011). Mandarin Broca aphasia and conduction aphasia usually involve damage to the left anterior insula (Mazzocchi & Vignolo, 1979; Mohr et al., 1978). These lesions affect

production of T3 and T2 more than T1 and T4 (Shi & Li, 2011). The current results showed that the left anterior insula is not only involved in laryngeal muscle control, but is also specifically involved in control of fundamental frequency rises (and this observation appears to be consistent with the neurological data on Mandarin speakers).

When the group level statistical threshold was lowered, the right putamen also showed stronger brain activation during T2 production than during T4 production. This finding is consistent with previous evidence that the putamen is involved in motor sequence organization by providing internal timing cues (Cunnington, Bradshaw, & Iansek, 1996; McFarland & Haber, 2002). Putaminal lesions cause dysarthria and dysphonia in humans but have no effect on monkey vocalizations (Jürgens, 2002), which suggests that the putamen is only involved in learned voluntary voice and speech production, not in the production of innate vocalizations. Thus, the stronger activation of the putamen in T2 suggests that T2 production needs more learned voluntary control of laryngeal movement, which is consistent with the fact that T2 is acquired later than T4.

4.1.2. Validation of the regionally stronger neural activation in T2 than in T4

It has been hypothesized that T1 involves setting fundamental frequency level, but does not involve rise or fall activity. Consequently, the comparison between T2 and T1 should leave brain activations that are specific to tone rise control unaffected. The results revealed stronger brain activation of the bilateral insula in T2 than in T1, which confirmed the involvement of the insula in tone rise control. However, there were some potentially important differences between the T2–T4 and T2–T1 comparisons: T2–T4 showed differences in the anterior insula in the left hemisphere, whereas T2–T1 showed differences in the left posterior insula. Previous evidence has shown that the anterior and posterior parts of the insula may have different functions: While the anterior insula is primarily involved in articulation coordination, the posterior insula may be more involved in somatic control (Kurth et al., 2010; Stephani, Fernandez-Baca Vaca, Maciunas, Koubeissi, & Luders, 2011). Thus, T1 may require the same level of laryngeal muscle coordination as T2 to sustain the muscle contraction at a stable level over a period of time whereas T2 may need more somatic information than T1.

Another validation of the site of neural activation in T2 came from the comparison between T2 and T3. Although T3 involves both tone rise and fall, the age of acquisition and neurological evidence suggested that T3 operates similarly to T2. Thus, it was hypothesized that comparison between T3 and T2 would cancel most of the brain activations that are involved in fundamental frequency rise control. The results confirmed this hypothesis since neither the insula nor the putamen showed any differences between T3 and T2. The only weaker brain activation in T3 than in T2 was located in the right cerebellum (VIII). This is a region where activity was not observed for either T2 or T4 (see below). The extensive cancellation of activity associated with T2 further validated the conclusion that the insula and putamen are specifically associated with tone rise control.

4.2. Neural mechanism for tone fall

4.2.1. Regionally stronger neural activation in T4 than in T2

A region located at the boundary of the left brainstem showed stronger activation during T4 than during T2. Because the anatomical structure in the cerebellum and brainstem are complex, the resolution of the fMRI procedure employed could not locate the position of the activation in these regions precisely. For this reason, only an approximate location of activation that is near the brainstem and cerebellum can be given. This region is close to

the substantia nigra, pontine reticular formation, and deep cerebellar nuclei, but not the periaqueductal gray matter (Bear, Connors, & Paradiso, 2007, pp. 224–225). Locating the activation in the region indicated would be consistent with animal and neuroanatomical evidence that the periaqueductal gray matter is not involved in voluntary voice control (Jürgens, 2002). Evidence showed that the brain stem reticular formation, specifically in its dorsal and parvocellular reticular nuclei have a close relationship with the motor control of voice (Bernard, Villanueva, Carroue, & Le Bars, 1990; Thoms & Jurgens, 1987; VanderHorst, Terasawa, & Ralston, 2001). These regions are involved in vocal motor coordination of both innate and learned voice production (Jürgens & Ehrenreich, 2007). The current results suggested that these regions were specifically involved in tone fall control.

4.2.2. Validation of the regionally stronger neural activation in T4 than in T2

The left brainstem activity remained when T4 and T1 were compared, so this region appears to be involved in fall control. Previous literature indicated that T1 and T4 are usually acquired earlier than T2 and T3. Based on this, it was assumed that T1 would be more similar to T4 than T2 with regards to the neural mechanisms involved. However, our results showed differences in both the left brainstem and in the temporal associative auditory cortex. The temporal associative auditory cortex has been implicated in auditory feedback control during speech production (Guenther, Ghosh, & Tourville, 2006). Thus, the additional involvement of the motor and auditory cortex in T4 as compared with T1 suggested that the neural mechanism in T4 is not simply a cessation of activity in the neural mechanisms that leads to rises. Previous evidence showed that singers relied more on auditory feedback to control fundamental frequency than did nonsingers (Jones & Keough, 2008). While fundamental frequency is processed in the auditory cortex (Hall, Edmondson-Jones, & Fridriksson, 2006), the additional involvement of auditory cortex in T4 suggests more complex neural mechanisms are involved in its control.

4.3. The contrast between the neural mechanisms for tone rise and fall

The SEM results showed that the model (reciprocal connection between the LMC and insula, unidirectional projection from the LMC to the putamen and brainstem) achieved a good match with the data of both T2 and T4. This finding is consistent with previous animal and neuroimaging evidence that shows that the LMC and its connection with other cortical and subcortical brain regions constitute a common neural network that is responsible for laryngeal control (Simonyan & Horwitz, 2011). Furthermore, the projection from the LMC to the brainstem showed a significant positive connection during T2 production, but did not reach significance in T4 production. As discussed in the introduction, physiological evidence indicated that in tone languages and singing an increase in the activity level of the cricothyroid muscle precedes rises in fundamental frequency (Lindestad, Fritzell, & Persson, 1991; Liu, Behroozmand, Bove, & Larson, 2011). Also, the suprahyoid muscles increase fundamental frequency (Ohala, 1972; Zenker, 1964). Conversely, when fundamental frequency falls from high to low, the cricothyroid muscle relaxes and the strap muscles show activity (Erickson, Baer, & Harris, 1983; Liu et al., 2011; Roubeau, Chevie-Muller, & Saint Guily, 1997).

Control of different laryngeal muscles is required in rises and falls, which makes it likely that the LMC to brainstem pathway is involved in both tone rises and falls. During fundamental frequency rises, the LMC projects information to the brainstem to increase the contraction of the cricothyroid, whereas during fundamental frequency falls, that same projection may disappear so that the cricothyroid is relaxed when T4 is produced. In this

case, a single neural mechanism that operates in different ways appears to exist for achieving fundamental frequency rise versus fall. This conclusion is consistent with the fact that T4 also required an increase in vocal fold tension to prepare for the fall in fundamental frequency at the onset of T4 that is not audible. Such increases are not seen in the idealized fundamental frequency contours shown in Fig. 1. More realistic patterns can be seen in Fig. 2 of Xu (1997), where the initial rise in T4 is apparent. Xu (1997) also shows that an initial fall occurs in T2 and this is approximately the same magnitude as the rise at the beginning of T4. Whereas the earlier account for the initial rise on T4 suggests that it is preparatory for the fall, Xu and Wang (2001) consider that this and the initial fall at the start of T2 is part of the tonal execution which occurs exclusively within the same syllable.

The SEM results also showed significant differences between T2 and T4 in the connection of the insula to the LMC. Specifically, there was a significant negative and a significant positive connectivity from the left anterior insula to the left LMC in T2 and T4, respectively. The involvement of this connection in fundamental frequency control is consistent with previous evidence that suggests that a reciprocal connection between the LMC and the insula is implicated in motor preparation and processing of all components of speech production (Greenlee et al., 2007; Simonyan et al., 2009). However, the significant positive and negative connection of the insula and LMC seems to be contrary to the above mentioned single passive mechanism hypothesis. Furthermore, the rise and fall of fundamental frequency may also have involved increased or decreased subglottal air pressure (Herman et al., 1996; Mosen et al., 1978). Brain regions involved in respiratory control, especially the inferolateral sensory region and auditory cortex associated with expiration, may also be involved in fundamental frequency control (Ludlow, 2005). Thus, it is possible that multiple neural mechanisms are involved.

McNamara et al. (2008) found that brain activity showed negative correlation with behavioral performance during sound–action association learning. It is possible that a similar negative correlation occurs between the insula and the LMC: When the LMC can make the control efficiently, less information input would be required from the insula; When the LMC requires additional information to perform the control, such as to stop projecting information to the brainstem so as to produce a fall, there would be a positive connection between the insula and the LMC. The problem with this explanation is why was the connection of the insula to the LMC significantly negative rather than non-significant or positive. One possible explanation may be that the stoppage of information input from the insula to the LMC is controlled by the insula, not the LMC. In this case, the insula would need to actively inhibit its input to the LMC. Some support for this explanation is the significantly stronger activation in the insula in T2 than in T4/T1. In sum, from the perspective of the whole network for fundamental frequency control, it seems that different neural mechanisms were involved in rise and fall.

4.4. Further explanation about the similarity and difference between T2 and T4

The results showed that although there is similarity in T2 and T4 control, the neural control of T2 involved more complicated mechanisms than those used in control of T4. This would result in later acquisition of T2 as compared with T4 (Li & Thompson, 1977) and would be consistent with the tone confusion literature (Clumeck, 1977; Li & Thompson, 1977). Also, a study on Indonesian Adult Students' acquisition of Mandarin tones showed that the acquisition of T1 and T4 was achieved quicker than T2 and T3 and that during learning T2 is often produced as a tone similar to T1, and T3 as a tone similar to T2 (Wang, 2006). However, other studies show that

people with different native languages acquire tones similar to T1–T4 in different orders from those in Mandarin (Yi & Liang, 2010). Recent neuroimaging evidence showed that language-dependent enhancement of pitch representation can transfer to other languages with similar phonological systems (Krishnan et al., 2010). Thus, native language experience may have an important influence on the neural mechanisms used in tone control.

4.5. Limitations and future work

First, the fact that the SEM connectivity analysis came out clearly with a design that involved extraction of one tone from sequences that included the other three, points to the robustness of the results. However, there are other issues associated with use of SEM that will be addressed in future work. For instance alternative approaches such as dynamic causal modeling should be used to quantify differences in effective connectivity rather than SEM. Second, the neural pattern of T3–T4 showed activation in the left insula that was similar to that observed with T2. The additional involvement of the right middle frontal region and temporal regions in T3 further indicated that T3 is not identical with T4, nor is T3 a simple addition of T2 and T4. Future studies are needed to further clarify the neural mechanisms behind T3 control. Third, if native language experience does have an effect on the neural control of tone, direct comparison of English speakers or speakers of other non-tonal languages acquiring the tones, and Mandarin speakers who have acquired the tones, would be helpful in further elucidating the neural mechanism for tone production. Fourth, the current study did not include a non-linguistic or non-tonal laryngeal musculature control condition. Such control conditions employing other types of laryngeal control could have been helpful in clarifying the extent of differential somatotopy in the motor cortex that can be resolved with fMRI. Future study would include such control conditions. Fifth, an explicit limitation of the current study is that when the participants continuously heard scanner noise, their productions may have been masked causing potential changes in production such as the Lombard effect. Although during the experiment the participants wore headphones that effectively reduced the background noise, influences of speaking environment should be examined in future work. Finally, it should be noted that the actual differences between T2 and T4 are subtle. Although this may not be surprising given the gestures are over-learned, the conclusion should be dealt with cautiously until they are replicated.

5. Conclusion

Relatively little is known about the neural control of tone rise and fall. Different hypotheses have been offered in which either single or multiple neural control mechanisms have been proposed. The present results showed that the left insula and right putamen were involved in tone rise control, whereas the left brainstem was involved in tone fall control. The results were validated by comparison of different pairs of tone which make different neural demands on the brain. The connectivity results further showed that the significant projection from the LMC to the brainstem used in T2 ceased its activity in T4. This supported the single neural control mechanism hypothesis. Additionally, the cessation of activity that controls T2 may have resulted from a different connection from the left insula to the LMC in T4 than that in T2, supported the multiple neural control mechanisms hypothesis.

Acknowledgment

This work was supported by a Grant from the National Natural Science Foundation of China (NSFC) (30900393).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bandl.2012.01.004.

References

- Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*, 89(2), 320–328.
- Ackermann, H., & Riecker, A. (2010). The contribution(s) of the insula to speech production: A review of the clinical and functional imaging literature. *Brain Structure and Function*, 214(5–6), 419–433.
- Bear, M., Connors, B., & Paradiso, M. (2007). *Neuroscience. Exploring the brain*. Lippincott Williams & Wilkins.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Bernard, J. F., Villanueva, L., Carroue, J., & Le Bars, D. (1990). Efferent projections from the subnucleus reticularis dorsalis (SRD): A Phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience Letters*, 116(3), 257–262.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, 32(2), 821–841.
- Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, 18(4), 837–845.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications.
- Büchel, C., Coull, J. T., & Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science*, 283(5407), 1538–1541.
- Clumek, H. V. (1977). *Studies in the acquisition of Mandarin phonology. Unpublished doctoral dissertation*. University of California, Berkeley.
- Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3000 monosyllabic words. *Behavior Research Methods*, 40(3), 791–794.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162–173.
- Cunnington, R., Bradshaw, J. L., & Iansek, R. (1996). The role of the supplementary motor area in the control of voluntary movement. *Human Movement Science*, 15, 627–647.
- Erickson, D. (1993). Laryngeal muscle activity in connection with Thai tones. *Research Institute of Logopedics and Phoniatrics Annual Bulletin*, 27, 135–149.
- Erickson, D., Baer, T., & Harris, K. S. (1983). *The role of the strap muscles in pitch lowering. Vocal Fold Physiology*. San Diego: College-Hill press.
- Evans, K. C., Shea, S. A., & Saykin, A. J. (1999). Functional MRI localisation of central nervous system regions associated with volitional inspiration in humans. *The Journal of Physiology*, 520(2), 383–392.
- Fiebach, C. J., Friederici, A. D., Smith, E. E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual–semantic representations in verbal working memory. *Journal of Cognitive Neuroscience*, 19(12), 2035–2049.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33(5), 636–647.
- Gandour, J., & Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language*, 18(1), 98–114.
- Gandour, J., Petty, S. H., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language*, 35(2), 201–240.
- Gandour, J., Ponglorpisit, S., Khunadorn, F., Dehongkit, S., Boongird, P., & Boonklam, R. (1992). Timing characteristics of speech after brain damage: Vowel length in Thai. *Brain and Language*, 42(3), 337–345.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4), 870–878.
- Greenlee, J. D. W., Oya, H., Kawasaki, H., Volkov, I. O., Severson, M. A., Howard, M. A., et al. (2007). Functional connections within the human inferior frontal gyrus. *Journal of Comparative Neurology*, 503(4), 550–559.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.
- Hall, D. A., Edmondson-Jones, A. M., & Fridriksson, J. (2006). Periodicity and frequency coding in human auditory cortex. *European Journal of Neuroscience*, 24(12), 3601–3610.
- Hannig, S., & Jürgens, U. (2006). Projections of the ventrolateral pontine vocalization area in the squirrel monkey. *Experimental Brain Research*, 169(1), 92–105.
- Harris, K. S. (1974). Physiological aspects of articulatory behavior. In T. A. Sebeok (Ed.), *Current trends in linguistics* (pp. 2281–2302). The Hague: Mouton.
- Harvey, N., & Howell, P. (1980). Isotonic vocalis contraction as a means of producing rapid decreases in Fo. *Journal of Speech and Hearing Research*, 23(3), 576–592.
- Herman, R., Beckman, M., & Honda, K. (1996). *Subglottal pressure and final lowering in English*.
- Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. *Folia Phoniatrica*, 22(1), 1–20.
- Honey, G. D., Suckling, J., Zelaya, F., Long, C., Routledge, C., Jackson, S., et al. (2003). Dopaminergic drug effects on physiological connectivity in a human cortico-striato-thalamic system. *Brain*, 126(Pt 8), 1767–1781.
- Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3–42.
- Institute of Linguistics, B. L. C. (1986). *A Frequency Dictionary for Modern Chinese* (p. 1491). Beijing: Beijing Language College Press.
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, 190(3), 279–287.
- Jöreskog, K. G. (1996). Modeling development: Using covariance structure models in longitudinal research. *European Child and Adolescent Psychiatry*, 5(Suppl 1), 8–10.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26(2), 235–258.
- Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research*, 1148, 90–95.
- Kast, M., Bezzola, L., Jancke, L., & Meyer, M. (2011). Multi- and unisensory decoding of words and nonwords result in differential brain responses in dyslexic and nondyslexic adults. *Brain and Language*, 119(3), 136–148.
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, 23(1), 81–95.
- Kuna, S. T., Smickley, J. S., Vanoye, C. R., & McMillan, T. H. (1994). Cricothyroid muscle activity during sleep in normal adult humans. *Journal of Applied Physiology*, 76(6), 2326–2332.
- Kurth, F., Eickhoff, S. B., Schleicher, A., Hoemke, L., Zilles, K., & Amunts, K. (2010). Cytoarchitecture and probabilistic maps of the human posterior insular cortex. *Cerebral Cortex*, 20(6), 1448–1461.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in mandarin-speaking children. *Journal of Child Language*, 4(2), 185–199.
- Lin, H. (2001). *A grammar of mandarin Chinese*. Muenchen: Lincom Europa.
- Lindestad, P. A., Fritzell, B., & Persson, A. (1991). Quantitative analysis of laryngeal EMG in normal subjects. *Acta Oto-Laryngologica*, 111(6), 1146–1152.
- Liu, H., Behroozmand, R., Bove, M., & Larson, C. R. (2011). Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback. *Journal of the Acoustical Society of America*, 129(6), 3946–3954.
- Loucks, T. M., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *NeuroImage*, 36(1), 131–143.
- Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, 147(2–3), 205–222.
- MacNeilage, P. F. (1972). *Speech physiology*. New York: Academic Press.
- Mazzocchi, F., & Vignolo, L. A. (1979). Localisation of lesions in aphasia: Clinical-CT scan correlations in stroke patients. *Cortex*, 15(4), 627–653.
- McFarland, N. R., & Haber, S. N. (2002). Thalamic relay nuclei of the basal ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. *Journal of Neuroscience*, 22(18), 8117–8132.
- McIntosh, A. R., & Gonzalez-Lima, F. (1994a). Network interactions among limbic cortices, basal forebrain, and cerebellum differentiate a tone conditioned as a pavlovian excitator or inhibitor: Fluorodeoxyglucose mapping and covariance structural modeling. *Journal of Neurophysiology*, 72(4), 1717–1733.
- McIntosh, A. R., & Gonzalez-Lima, F. (1994b). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2, 2–22.
- McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. *Journal of Neuroscience*, 14(2), 655–666.
- McNamara, A., Buccino, G., Menz, M. M., Glascher, J., Wolbers, T., Baumgartner, A., et al. (2008). Neural dynamics of learning sound–action associations. *PLoS ONE*, 3(12), e3845.
- Mohr, J. P., Pessin, M. S., Finkelshtein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathologic and clinical. *Neurology*, 28(4), 311–324.
- Monsen, R. B., Engebretson, A. M., & Vemula, N. R. (1978). Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of fundamental frequency in English. *The Journal of the Acoustical Society of America*, 64, 65.
- Ohala, J. J. (1970). *Aspects of the control and production of speech*. UCLA Working Papers in Phonetics, 15.
- Ohala, J. J. (1972). How is pitch lowered. *Journal of the Acoustical Society of America*, 52, 124.
- Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey*. New York: Academic Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Olthoff, A., Baudewig, J., Kruse, E., & Dechent, P. (2008). Cortical sensorimotor control in vocalization: A functional magnetic resonance imaging study. *Laryngoscope*, 118(11), 2091–2096.
- Packard, J. L. (1986). Tone production deficits in nonfluent aphasic Chinese speech. *Brain and Language*, 29(2), 212–223.
- Peck, K. K., Galgano, J. F., Branski, R. C., Bogomolny, D., Ho, M., Holodny, A. I., et al. (2009). Event-related functional MRI investigation of vocal pitch variation. *NeuroImage*, 44(1), 175–181.
- Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, A. (2009). Auditory-motor integration during fast repetition: The neuronal correlates of shadowing. *NeuroImage*, 47(1), 392–402.

- Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. L. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Oto-Laryngologica*, 117(3), 459–464.
- Shi, Y., & Li, S. L. (2011). Mechanisms and performance of voice disorders in aphasia (review). *Chinese Journal of Rehabilitation Theory & Practice*, 17(2), 148–150.
- Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, 17(2), 197–208.
- Simonyan, K., & Jürgens, U. (2002). Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Research*, 949(1–2), 23–31.
- Simonyan, K., & Jürgens, U. (2005). Afferent cortical connections of the motor cortical larynx area in the rhesus monkey. *Neuroscience*, 130(1), 133–149.
- Simonyan, K., & Ludlow, C. L. (2011). Abnormal structure–function relationship in spasmodic dysphonia. *Cerebral Cortex*.
- Simonyan, K., Ostuni, J., Ludlow, C. L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *Journal of Neuroscience*, 29(47), 14912–14923.
- Stephani, C., Fernandez-Baca Vaca, G., Maciunas, R., Koubeissi, M., & Luders, H. O. (2011). Functional neuroanatomy of the insular lobe. *Brain Structure and Function*, 216(2), 137–149.
- Thoms, G., & Jurgens, U. (1987). Common input of the cranial motor nuclei involved in phonation in squirrel monkey. *Experimental Neurology*, 95(1), 85–99.
- Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice*, 3(3), 213–224.
- VanderHorst, V. G., Terasawa, E., & Ralston, H. J. 3rd., (2001). Monosynaptic projections from the nucleus retroambiguus region to laryngeal motoneurons in the rhesus monkey. *Neuroscience*, 107(1), 117–125.
- Wang, M. (2006). Indonesian Chinese students' acquisition of Chinese tones. *Journal of College of Chinese Language and Culture of Jinan University*, 2, 10–31.
- Wang, J. (2004). Tone perception and production deficit of Chinese aphasia: A case study. *Chinese Journal of Physical Medicine and Rehabilitation*, 26(3), 146–147.
- Xiong, J., Gao, J. H., Lancaster, J. L., & Fox, P. T. (1995). Clustered pixels analysis for functional MRI activation studies of the human brain. *Human Brain Mapping*, 3, 287–301.
- Xu, Y. (1997). Contextual tonal variations in mandarin. *Journal of Phonetics*, 25, 61–84.
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from mandarin Chinese. *Speech communication*, 33(4), 319–337.
- Yi, B., & Liang, J. (2010). The research of Chinese tone acquisition as second language. *Journal of Tianjin Normal University (Social Science)*, 2.
- Yip, M. (2002). *Tone. Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.
- Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, 48(2), 607–618.
- Zenker, W. (1964). Questions regarding the function of external laryngeal muscles. In D. W. Brewer (Ed.), *Research potentials in voice physiology*. Syracuse, NY: State university of NY.
- Zhang, Q., & Yang, Y. (2003). The determiners of picture-naming latency. *Acta Psychologica Sinica*, 35(4), 447–454.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8), 1770–1781.