Specificity of experience-dependent pitch representation in the brainstem

Yisheng Xu[a], Ananthanarayan Krishnan[b] and Jackson T. Gandour[b]

[a]Center for Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania and [b]Department of Speech, Language, & Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

Correspondence and requests for reprints to Professor Jackson T. Gandour, PhD, Purdue University, Department of Speech, Language, & Hearing Sciences, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette 47907-2038, IN, USA
Tel: + 1 765 494 3821; fax: + 1 765 494 0771; e-mail: gandour@purdue.edu

This article is based on part of a doctoral dissertation completed by the first author at Purdue University in December 2005. Y.X. is currently a postdoctoral trainee in the Center for Neural Basis of Cognition at Carnegie Mellon University.

Sponsorship: Purdue Research Foundation dissertation grant (J.G.); NIH research grant R01 DC04584-05 (J.G.).

Received 19 July 2006; accepted 24 July 2006

Crosslanguage comparisons of brainstem-evoked potentials have revealed experience-dependent plasticity in pitch representation for curvilinear f_0 contours representative of Mandarin tones. To assess the tolerance limits of this experience-dependent selectivity, we evaluated cross-linguistically (Chinese, English) the pitch strength and tracking accuracy of linear rising and falling f_0 ramps representative of Mandarin tones 2 and 4. No crosslanguage differences in pitch strength or accuracy were observed for either tone, indicating that stimuli with linear rising/falling ramps elicit homogeneous pitch representations at the level of the brainstem regardless of language experience. We conclude that pitch extraction at the brainstem level is critically dependent on specific dimensions of pitch contours that native speakers have been exposed to in natural speech contexts. *NeuroReport* 17:1601–1605 © 2006 Lippincott Williams & Wilkins.

Keywords: auditory, brainstem, experience dependent plasticity, frequency following response, lexical tone, Mandarin Chinese, pitch

Introduction

Regardless of the neural mechanisms underlying lower-level processing of spectral and temporal cues [1], hemispheric specialization is clearly sensitive to higher order information about the linguistic status of the auditory signal [2]. Whereas neural specializations are indeed predictable on the basis of low-level features of the stimulus (cue-specific), they can also be influenced by linguistic status (domain-specific). In the case of lexical tones, crosslanguage studies have revealed activation in left hemisphere regions for native speakers of tone languages only [3]. These data notwithstanding, hemispheric specialization at the cortical level cannot be fully accounted for by claiming one (cue-specific) or the other (domain-specific) as the sole explanatory model [2]. Besides the cortical level, a complete understanding of the neural organization of language can only be achieved by viewing language processes as a set of computations or mappings between representations at different stages of processing [4]. Language-dependent operations may begin before the signal reaches the cerebral cortex. The degree of linguistic specificity is yet to be determined for computations related to pitch representation at the level of the auditory brainstem.

Previous research has demonstrated that selective attention to nonspeech periodic tones may modify short-latency auditory-evoked responses in the rostral brainstem, as measured by the frequency following response (FFR) [5]. Using English vowels (female /a/, male /e/) in a dichotic listening task, FFRs associated with voice fundamental frequency (f_0) are larger in the attended condition than in the unattended condition [6], suggesting that brainstem neurons are sensitive to attention to paralinguistic information (e.g. sex of speaker). Moreover, brainstem neurons appear to be sensitive to linguistic information. A comparison of forward and reverse speech stimuli reveals that FFRs show increased amplitude in forward speech [7]. This finding suggests that familiar phonological and prosodic properties of forward speech may affect short-latency brainstem-evoked potentials.

A recent study of such prosodic properties that occur in natural speech shows that pitch processing at the brainstem level may be influenced by language experience [8]. Specifically, the pitch strength and accuracy of pitch tracking of Mandarin Chinese tones are significantly greater in native listeners than in nonnative listeners. These data suggest that neural plasticity at the brainstem level may be induced by language experience that enhances or primes linguistically relevant features of the speech input. In the study by Krishnan *et al.* [8], however, stimuli exhibited prototypical, curvilinear f_0 contours that were modeled after Mandarin tones in natural speech [9]. If brainstem reorganization is induced by speech-specific experience, the question arises as to what the tolerance limits are for linguistic sensitivity at this subcortical level. What specific f_0 properties or features of the stimulus, static or dynamic, are relevant? To what extent can a stimulus deviate from natural speech exemplars before exceeding the upper or the lower limit of linguistic sensitivity of brainstem neurons?

Behavioral data have shown that even nonprototypical tonal contours may induce crosslanguage differences in speech perception. In multidimensional scaling of listeners' perception of linear f_0 ramps (level, falling, rising, falling–rising, rising–falling), crosslanguage comparisons show that the relative importance of the pitch height and direction (rising vs. nonrising) dimensions varies depending on a listener's familiarity with specific types of pitch patterns that occur in the tone space of their native language [10,11]. Using a linear f_0 continuum ranging from a Mandarin tone 2 (high-rising) to a tone 1 (high level), crosslanguage (Mandarin, English) comparisons reveal that categorical perception for pitch direction is dependent on a listener's experience with a tone language [12].

These behavioral data notwithstanding, the aim of this experiment is to determine whether linear f_0 ramps, similar to Mandarin tone 2 (high rising) and tone 4 (high falling) in direction but dissimilar in trajectory, elicit brainstem FFRs differentially as a function of language experience (Chinese, English). By including a nontone language group (English), we can evaluate whether any observed effects on pitch representations are language-universal irrespective of experience with lexical tones. By comparing FFRs elicited by a linear vs. curvilinear [8] model of tone 2, we can begin to assess the tolerance limits for priming linguistically relevant features of the auditory signal useful for pitch extraction at the level of the brainstem. The absence of a language experience effect in response to the linear tone 2, coupled with our earlier findings for the curvilinear tone 2 [8], would be consistent with the notion that neural mechanisms underlying experience-dependent selectivity are local to the generators of the FFR in the human brainstem. Finally, a comparison of FFRs elicited by a linear model of tone 2 vs. tone 4 permits us to evaluate the degree to which pitch direction (rising vs. falling) interacts with f_0 linearity at the level of the brainstem.

Materials and methods

Participants

Sixteen native speakers of Mandarin (8 men; 8 women) and 16 native speakers of American English (7 men; 9 women) participated in the study. The two groups were closely matched in age (mean ± SD: Chinese=27.8 ± 2.9; English=25.5 ± 3.7) and years of formal education (mean ± SD: Chinese=18.8 ± 2.6; English=18.1 ± 2.7). All participants were right-handed. Hearing sensitivity was better than 20 dB hearing level (HL) for octave frequencies from 500 to 4000 Hz. All Chinese participants were from mainland China and none had received any formal instruction in English until after the age of 11 years. English participants had no previous exposure to Chinese or any other tone language. None of the participants, Chinese or English, had more than 5 years of formal musical training, and none had any musical training within the past 5 years. Participants were paid for taking part in the study and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

Stimuli

A Mandarin monosyllable [i] with linear rising and falling f_0 ramps was created using a synthesis-by-rule scheme [13] (Fig. 1). The ramps for the rising (tone 2) and falling (tone 4) contours ranged from 90 to 140 Hz and from 140 to 90 Hz, respectively. Duration was fixed at 250 ms. Amplitude was constant at 60 dB. Vowel formant frequencies were steady-state (in Hz): F_1=270; F_2=2290; F_3=3010; and F_4=4000 [14]. The two synthetic speech sounds were judged by five native listeners to be good quality Mandarin words *yi*^{2} ('aunt') and *yi*^{4} ('easy').

Experimental protocol

Participants reclined comfortably in an acoustically and electrically shielded booth. All stimuli were controlled by a signal generation and data acquisition system (System III, Tucker-Davis Technologies, Gainesville, Florida, USA). The stimulus files were routed through a digital to analog module, and presented binaurally to both ears through magnetically shielded insert earphones (TIP-300, Biologic, Mundelein, Illinois, USA). All stimuli were presented at 80 dB sound pressure level with a repetition rate of 2.1/s. The order of the two synthetic speech sounds was randomized across participants.

FFRs were recorded from active electrodes on the midline of the forehead at the hairline referenced to the linked mastoids. Another electrode placed on the mid-forehead (Fpz) served as the common ground. The interelectrode impedances were maintained below 3000 Ω. The electroencephalogram inputs were amplified by 200 000 and bandpass filtered from 100 to 3000 Hz (6 dB/octave roll-off, response characteristics). Each FFR waveform represents an average of 1500 stimulus presentations over a 0.3-s analysis window using a sampling rate of 50 kHz.

Data analysis

Each FFR waveform was cross-correlated with the corresponding stimulus waveform to estimate its latency [15]. A time-lag window between 0.002 and 0.015 s (based on the latency range of FFR [16]) was selected to locate the maximum cross-correlation peak and to compute its time lag (latency). The leading portion of the FFR in the duration of this time lag was trimmed to correct the intersubject variance of latencies. A trailing portion outside the stimulus duration was also trimmed, resulting in a 0.25-s time window for the subsequent analysis.

The ability of FFRs to follow the pitch change in the stimuli was evaluated by extracting the f_0 contour from the FFR time series using a periodicity detection short-term autocorrelation algorithm [17]. This analysis was performed on 22 successive small frames (0.01 s) from 0.02 to 0.23 s taken from the time series, yielding estimates of both pitch periodicity and pitch strength. Pitch periodicity is defined as the time lag associated with the maximum autocorrelation peak; pitch strength is defined as the autocorrelation peak coefficient ranging from 0 to 1. The autocorrelation algorithm provides a reliable measure of pitch strength or salience [8,18,19].

The pitch strength of each FFR waveform was derived by averaging the pitch strength across all the short-term autocorrelation frames. The f_0 contour of FFR was derived from the inverse of pitch periodicity of each successive frame. A cross-correlation between the FFR and

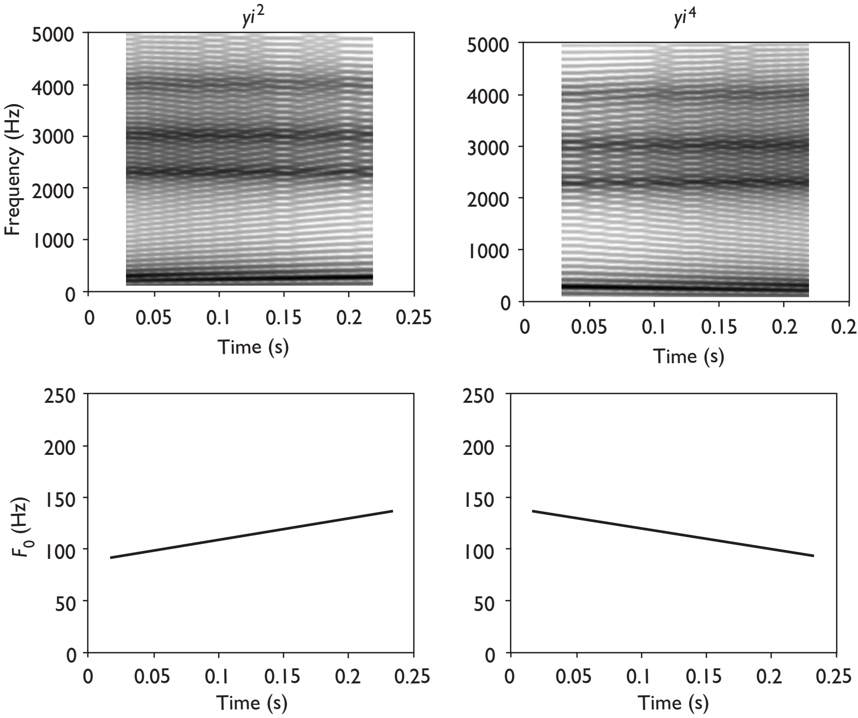

Fig. 1 Spectrograms (top panels) and f_0 contours (bottom panels) of Mandarin Chinese synthetic speech stimuli: yi^2 'aunt', rising linear ramp; yi^4 'easy', falling linear ramp.

corresponding stimulus f_0 contours was performed to estimate the goodness of fit between the two contours. Owing to a skewed distribution, cross-correlation coefficients were transformed to ranks in order to represent the relative order of pitch tracking accuracy among all the acquired FFR time series. Pitch strength and pitch tracking accuracy were then analyzed using mixed model analyses of variance (ANOVAs) with participants as a random effect to estimate the effect of language group (Chinese, English) and pitch direction (rising, falling).

Results

Pitch strength

The mean FFR pitch strength for each group and pitch direction are plotted in Fig. 2a. Results of the ANOVA revealed no significant main effect for either group [F(1,30)=0.90; *P*=0.3491] or pitch direction [F(1,30)=0.61; *P*=0.4421]. No interaction was seen between the language group and pitch direction [F(1,30)=1.06; *P*=0.3116].

Pitch tracking accuracy

The mean FFR pitch tracking accuracy for each group and pitch direction is plotted in Fig. 2b. The ANOVA results for pitch tracking accuracy similarly showed no significant effects for group [F(1,30)=0.17; *P*=0.6871] and pitch direction [F(1,30)=0.02; *P*=0.8812]. No interaction was seen between the language group and pitch direction [F(1,30)=0.51; *P*=0.4795].

Discussion

The major finding of this study is that nonnative (English) and native (Chinese) listeners are homogeneous in terms of degree of pitch strength and pitch tracking accuracy when responding to a synthetic speech stimulus with either a linear rising (tone 2) or falling (tone 4) f_0 ramp. These data are seen to complement, rather than conflict with, our earlier observation of crosslanguage differences in response to synthetic speech stimuli with curvilinear rising or falling f_0 contours [8]. That is, in contrast to a prototypical, curvilinear stimulus representative of tones 2 and 4, stimuli with a linear ascending/descending ramp may elicit homogeneous pitch representations at the level of the brainstem regardless of language experience.

Our explanation is that crosslanguage differences in pitch extraction at the brainstem level are critically dependent on the specific dimensions of pitch contours that native speakers are exposed to. In the case of Mandarin tones 2 and 4, native listeners' long-term learning experience has improved their ability to rapidly track nonlinear changes in pitch movement at the syllable level with a high degree of accuracy. No language-dependent effects, however, are observed in response to linear rising or falling f_0 ramps because they are not part of native Chinese listeners' experience. We therefore conclude that rising or falling f_0 movement alone is insufficient to induce changes in the pattern of neural responses at the brainstem level. At this subcortical level, neural mechanisms that mediate experience-dependent pitch representation appear to be acutely sensitive to dynamic changes in trajectory throughout the duration of a pitch contour. More generally, sustained phase-locked activity in the brainstem, representing pitch relevant information, is selectively 'tuned' to the specific interspike intervals that correspond to pitch contours from the listeners' experience. One possible encoding scheme is that the phase-locking at the pitch periods corresponding to the pitch contour is enhanced by both

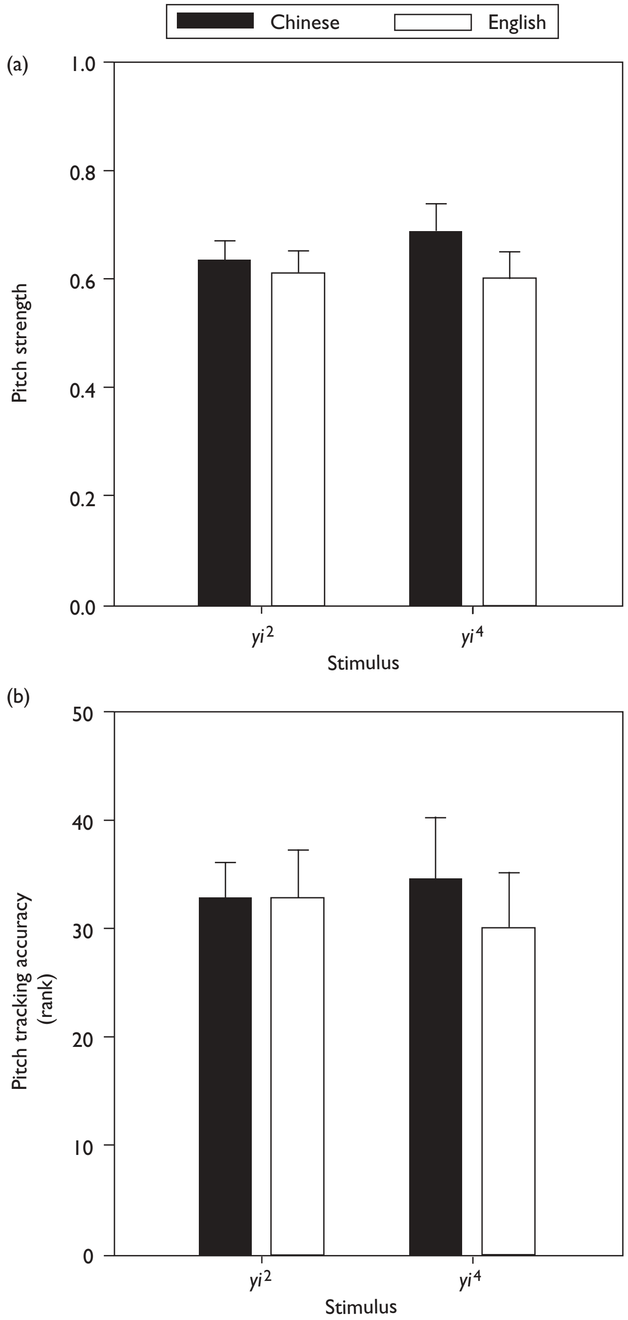

Fig. 2 Crosslanguage comparisons of pitch strength and pitch tracking accuracy for the rising (yi^2) and falling (yi^4) f_0 ramps. No language experience effects are evident for the linear ramps in either direction.

active facilitation/disinhibition and/or suppression of phase-locking at intervals other than the pitch period(s). Excitatory and inhibitory interactions are known to play an important role in signal selection at the level of the brainstem [20,21]. Thus, it appears that the reorganization in the brainstem for pitch extraction is specific to particular dimensions of the auditory stimuli that are dependent on a listener's experience. In this case, it is the curvilinear dimension of pitch contours that are manifest in Mandarin tones 2 and 4.

The absence of a language experience effect in response to the linear stimuli, as compared with the curvilinear stimuli [8], is consistent with the notion that neural mechanisms underlying experience-dependent selectivity are local to the generators of the FFR in the human brainstem. It is well known that the corticofugal system can lead to subcortical egocentric selection of behaviorally relevant stimulus parameters in nonprimate and nonhuman primate animals [22,23]. In the case of humans, although the corticofugal system likely facilitates the reorganization in the brainstem for pitch extraction in earlier stages of language development, it is unlikely that it continues to play a primary role in pitch extraction in the mature healthy adult brainstem. First, the corticofugal egocentric selection is short term and takes time (latency) to be activated whereas the FFR response latency is only about 6–8 ms. Second and a more compelling inference in favor of a local mechanism is the absence of a crosslanguage FFR effect at the brainstem level elicited by linear f_0 contours, in contrast to curvilinear contours [8]. In addition, multidimensional scaling data show that dimensions underlying the perception of linear f_0 contours, similar to those used in the current study, are weighted differentially as a function of language experience [10,11]. Cortical modulation of the brainstem response would have led us to expect, contrary to fact, differential pitch representations of the linear f_0 contours between Chinese and English listeners.

One difference in methodology is found between this study and the earlier one [8], which could be a potential confound in our interpretation of differences in experimental outcomes between the two studies. In this study, we employed binaural stimulation instead of monaural stimulation [8]. We chose binaural stimulation because no significant laterality effects were identified for stimuli presented monaurally to the left and the right ear in our previous studies [8,19]. Is it possible that binaural interaction among the brainstem neural elements differentially affects the pitch representation in such a way that crosslanguage differences are effectively washed out? We argue that this is very unlikely because binaural stimulation, as compared with the sum of independent monaural stimulations, typically produces an occlusive interaction in the brainstem yielding a smaller magnitude response instead of a response enhancement [24].

The findings of this study are compatible with the view that the tolerance limits for priming features useful for pitch extraction at the brainstem level are based on fine-grained phonetic features of tonal categories. The next question to address is exactly how fine-grained is this specificity for the reorganization of pitch encoding within the brainstem. For example, behavioral data indicate the presence of categorical boundaries for tonal continua in Mandarin [12]. Brain imaging data reveal a double dissociation between language group and native vs. nonnative tonal categories at the level of the auditory cortex [25]. Do these categorical effects influence processing at the level of the brainstem? Or are these limits predominantly set by lower-level physical characteristics of the stimuli such as the curvilinearity of pitch trajectories as demonstrated in the present experiment?

Conclusion

Experience-dependent enhancement of pitch representation at the brainstem level is specific to pitch contours within the native listeners' experience. Neural mechanisms mediating this experience-dependent enhancement are local to the generators of the FFR in the rostral brainstem.

References

1. Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 2003; **41**:245–255.
2. Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond Ser B Biol Sci* (in press).
3. Gandour J. Brain mapping of Chinese speech prosody. In: Li P, Tan L, Bates E, Tzeng O, editors. *Handbook of East Asian psycholinguistics: Chinese*. Cambridge, UK: Cambridge University Press; 2006. pp. 308–319.
4. Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 2004; **92**:67–99.
5. Galbraith GC, Olfman DM, Huffman TM. Selective attention affects human brain stem frequency-following response. *Neuroreport* 2003; **14**:735–738.
6. Galbraith GC, Bhuta SM, Choate AK, Kitahara JM, Mullen TA Jr. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 1998; **9**:1889–1893.
7. Galbraith GC, Amaya EM, de Rivera JM, Donan NM, Duong MT, Hsu JN, *et al*. Brain stem evoked response to forward and reversed speech in humans. *Neuroreport* 2004; **15**:2057–2060.
8. Krishnan A, Xu Y, Gandour J, Cariani PA. Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn Brain Res* 2005; **25**: 161–168.
9. Xu Y. Contextual tonal variations in Mandarin. *J Phonetics* 1997; **25**: 61–83.
10. Gandour J. Tone perception in Far Eastern languages. *J Phonetics* 1983; **11**:149–175.
11. Gandour J, Harshman RA. Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Lang Speech* 1978; **21**:1–33.
12. Xu Y, Gandour J, Francis A. Effects of language experience and stimulus complexity on categorical perception of pitch direction. *J Acoust Soc Am* 2006 (in press).
13. Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am* 1980; **67**:971–995.
14. Howie J. *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press; 1976.
15. Galbraith GC, Brown WS. Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. *Electroencephalogr Clin Neurophysiol* 1990; **77**:295–308.
16. Sohmer H, Pratt H, Kinarti R. Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 1977; **42**:656–664.
17. Boersma P, van Heuven V. Speak and unSpeak with PRAAT. *Glot Int* 2001; **5**:341–347.
18. Cariani PA, Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 1996; **76**:1698–1716.
19. Krishnan A, Xu Y, Gandour J, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004; **189**:1–12.
20. Ananthanarayan AK, Gerken GM. Post-stimulation effects on the auditory brain stem response partial-masking and enhancement. *Electroencephalogr Clin Neurophysiol* 1983; **55**:223–226.
21. Ananthanarayan AK, Gerken GM. Response enhancement and reduction of the auditory brain-stem response in a forward-masking paradigm. *Electroencephalogr Clin Neurophysiol* 1987; **66**:427–439.
22. Suga N, Ma X, Gao E, Sakai M, Chowdhury SA. Descending system and plasticity for auditory signal processing: neuroethological data for speech scientists. *Speech Commun* 2003; **41**:189–200.
23. Suga N, Gao E, Zhang Y, Ma X, Olsen JF. The corticofugal system for hearing: recent progress. *Proc Natl Acad Sci USA* 2000; **97**:11807–11814.
24. Krishnan A, McDaniel SS. Binaural interaction in the human frequency-following response: effects of interaural intensity difference. *Audiol Neurootol* 1998; **3**:291–299.
25. Xu Y, Gandour J, Talavage T, Wong D, Dzemidzic M, Tong Y, *et al*. Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum Brain Mapp* 2006; **27**:173–183.