# Experience-dependent neural representation of dynamic pitch in the brainstem

Ananthanarayan Krishnan, Jackson T. Gandour, Gavin M. Bidelman and Jayaganesh Swaminathan

**Brainstem frequency-following responses were recorded from Chinese and English participants in response to an iterated rippled noise homologue of Mandarin Tone 2 (T2) and linear and inverted curvilinear variants. Pitch-tracking accuracy and pitch strength analyses showed advantages for the Chinese group over the English in response to T2 only. Pitch strength was larger for the Chinese group in rapidly changing sections of T2 compared with corresponding sections of a linear ramp. We conclude that experience-dependent neural plasticity at subcortical levels of representation is highly sensitive to specific features of pitch patterns in one's native language. Such experience-dependent effects suggest that subcortical sensory encoding interacts with cognitive processing in the cerebral cortex to shape the perceptual system's response to pitch patterns.** *NeuroReport* **20:408–413** © **2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

NeuroReport 2009, **20**:408–413

**Keywords: auditory, cochlear implant, experience-dependent plasticity, fundamental frequency following response, human, iterated rippled noise, Mandarin Chinese, pitch**

Department of Speech Language Hearing Sciences, Purdue University, Indiana, USA

Correspondence to Professor Ananthanarayan Krishnan, PhD, Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA
Tel: +1 765 494 3793; fax: +1 765 494 0771; e-mail: rkrish@purdue.edu

**Received** 20 November 2008 **accepted** 5 December 2008

## Introduction

Language processes can be viewed as a set of transformations between representations at different stages of processing [1–3]. Early stages of subcortical auditory processing are influenced by experience and active training [4–9]. Pitch encoding in the auditory brainstem is shaped by language experience [10–12], suggesting that early processing stages may perform transformations on the acoustic data that are relevant to linguistic as well as nonlinguistic auditory perception.

The scalp-recorded human 'frequency following response' (FFR) reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem [13]. Iterated rippled noise (IRN) generates auditory stimuli that preserve the perception of pitch, but do not have waveform periodicity characteristic of speech stimuli [14,15]. Dynamic curvilinear IRN stimuli permit us to investigate neural mechanisms underlying pitch patterns representative of those that occur in natural speech without a semantic confound [12,16]. To assess the tolerance limits for priming linguistically relevant pitch, we first examined FFRs in response to linear ramps representative of Mandarin rising [Tone 2 (T2)] and falling tones in a speech context [17]. These linear ramps elicited homogeneous pitch representations at the level of the brainstem regardless of language experience.

The primary aim herein is to further examine the linguistic sensitivity of brainstem neurons in pitch encoding by employing 'linear' and 'curvilinear' IRN variants of Mandarin T2. In cross-language comparisons of FFRs elicited by trilinear and linear approximations of T2 versus its curvilinear exemplar, we can assess whether curvilinearity is necessary to prime linguistically relevant features of the auditory signal for pitch extraction at the level of the brainstem. By comparing FFRs elicited by an inverted curvilinear approximation of T2, a pitch pattern that is not native to the Mandarin tonal space, versus its curvilinear exemplar, we can assess whether curvilinearity itself is sufficient to trigger language-dependent effects.

## Methods

### Participants

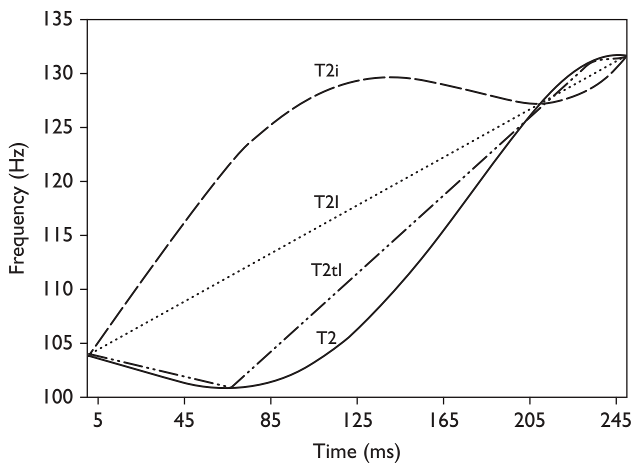
Ten adult native speakers of Mandarin Chinese and American English participated in the experiment. Participants' age ranged from 21 to 31 years. All Chinese participants were born and raised in mainland China. English participants had no previous exposure to a tone language. No participant had more than 2 years of musical training, and none had any musical training within the past 5 years. They gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

### Stimuli

A set of four pitch patterns, two curvilinear and two linear, was chosen to contrast an IRN homologue of a prototypical T2 (|yi$^2$| 'aunt') with three $f_0$ variants that do not occur in the Mandarin tonal space [18] (Fig. 1). The curvilinear variant (T2i) represented a polynomial

0959-4965 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e3283263000

**Fig. 1**




Iterated rippled noise homologues of time-normalized (250 ms) linear and curvilinear $f_0$ patterns. T2 (solid) is modelled after Mandarin Tone 2 using a fourth-order polynomial equation [18]; T2i (dash), inverted polynomial variant of T2; T2l (dot), linear variant of T2; T2tl (dash–dot–dot), trilinear variant of T2.

flip of T2. Of the two linear variants, one represented a linear ramp (T2l), preserving the onset and offset of T2; the other a trilinear approximation of T2 (T2tl), preserving the major points of inflection besides onset and offset. All time-varying IRN stimuli were created at a high iteration step ($n = 32$) using procedures described in Ref. [16]. Stimulus duration was 250 ms including a 10-ms cosine squared ramp used to eliminate spectral splatter and to minimize onset responses.

### Data acquisition

Data acquisition procedures were similar to those described in Refs [10,12,16]. FFRs were recorded from each participant in response to monaural stimulation of the right ear at 82 dB sound pressure level using a magnetically shielded insert earphone (Etymotic, ER3A; Elk Grove, Illinois, USA). These evoked responses were recorded differentially between a common noninverting (positive) electrode placed on the midline of the forehead at the hairline and inverting (reference) electrodes placed on (i) the ipsilateral mastoid, (ii) the contralateral mastoid, and (iii) the seventh cervical verterbra. Another electrode placed on the mid-forehead (Fpz) served as the common ground. FFRs were recorded simultaneously from the three different electrode configurations, and subsequently averaged for each stimulus condition to yield an FFR response with a higher signal-to-noise ratio. The interelectrode impedances were below 1 kΩ.

### Data analysis

Data analysis procedures were similar to those described in Refs [10,12].

#### *Temporal and spectral analysis*

Short-term autocorrelation functions and running autocorrelograms (ACGs) were computed from the averaged FFRs derived from each participant to index variation in FFR periodicities over the duration of the response (*c.f.* [10,16]). The ACG represents the short-term autocorrelation function of windowed frames of a compound signal. It is a three-dimentional plot quantifying periodicity and pitch strength variations over time. The horizontal axis represents time; the vertical axis represents the time lags associated with the peaks of the autocorrelation function and the intensity of each point in the image represents the amplitude of the autocorrelation function associated with a particular lag at a particular time. Narrow-band spectrograms were obtained from each FFR waveform to evaluate the spectral composition.

#### *Pitch-tracking accuracy*

The ability of the FFR to follow pitch change in the stimuli was evaluated by extracting the $f_0$ pattern from the averaged FFRs using a periodicity detection short-term autocorrelation algorithm [19]. The autocorrelation function was computed for each 40-ms frame after successive 10-ms shifts (*c.f.* [10]). The time lag corresponding to the maximum autocorrelation value for each frame was recorded for both the stimulus and FFR. The time lags associated with autocorrelation peaks in each frame were concatenated together to give a running $f_0$ contour. FFR pitch-tracking accuracy represents the cross-correlation coefficient between the $f_0$ contour extracted from the FFRs and the stimuli.

#### *Pitch strength*

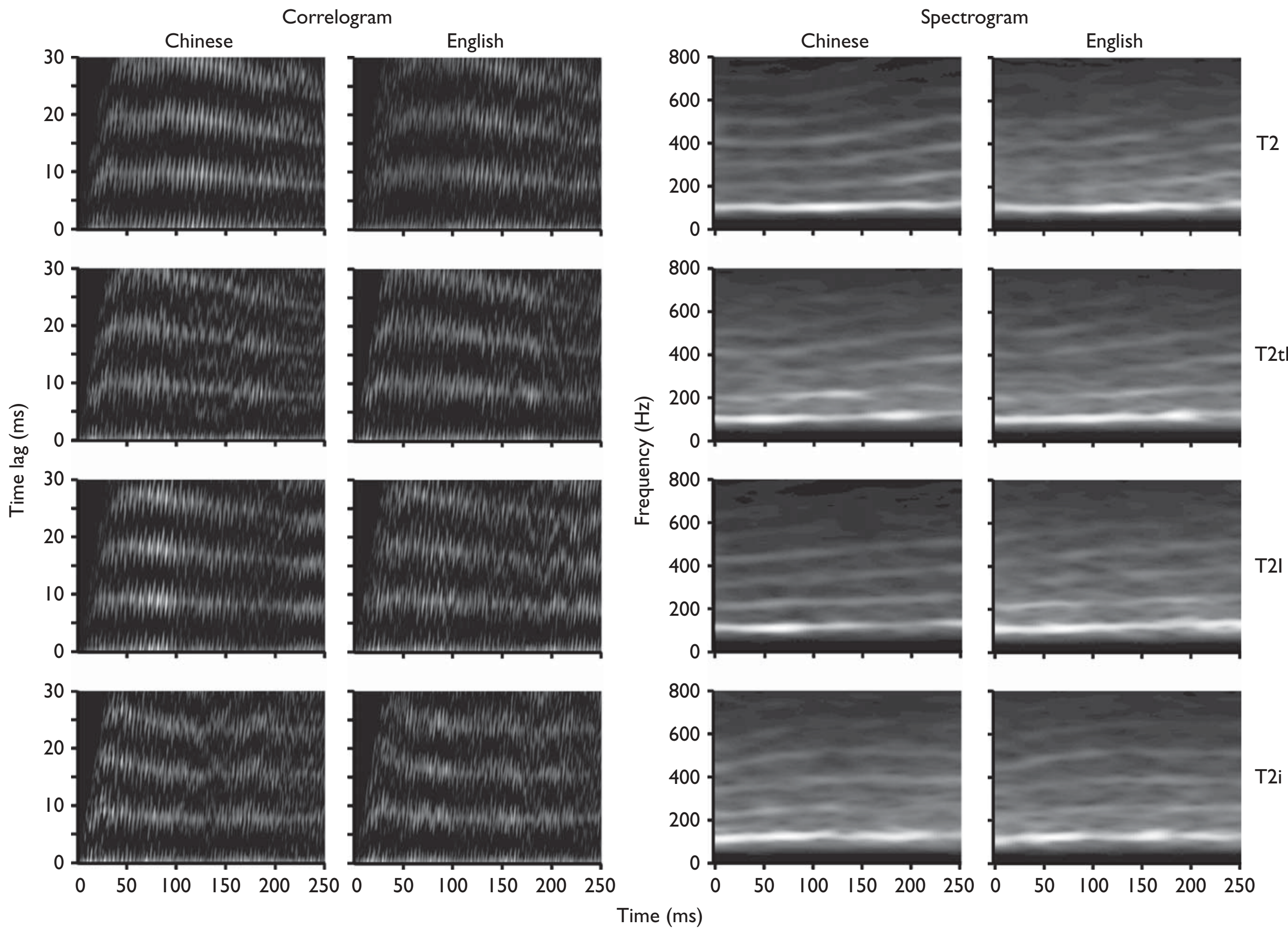
FFR responses were divided into six nonoverlapping 40 ms time frames (5–45; 45–85; 85–125; 125–165; 165–205; 205–245). The normalized autocorrelation function of the two language groups was derived from an analysis of corresponding time frames of the four stimuli and their FFR responses. Within each 40 ms frame, the response peak selected was the one that was closest to the location of the autocorrelation peak in the input stimulus [10,12]. This response peak was taken to be an estimate of pitch strength per time frame. Pitch strength was measured by the average magnitude of the normalized autocorrelation peak per language group in each 40-ms frame.

## Results

### Temporal and spectral properties

Grand averaged ACGs (left panels) and narrow-band spectrograms (right panels) are shown in Fig. 2 for the Chinese and English groups. In the Chinese group, ACGs show clear white bands of phase-locked activity at $f_0$ and its multiples in response to a native pitch pattern (T2), but less distinct and more diffuse bands in response to others, curvilinear and linear alike. In the English group, the bands are less distinct and more diffuse across the

**Fig. 2**




Grand averaged correlograms (columns 1 and 2) and spectrograms (columns 3 and 4) derived from grand averaged frequency following response (FFR) waveforms of Chinese and English groups in response to the four iterated rippled noise homologues of Mandarin Tone 2 and its variants (T2, row 1; T2tl, trilinear variant of T2, row 2; T2l, linear variant of T2, row 3; T2i, inverted polynomial variant of T2, row 4). In response to T2, correlograms of the Chinese group (column 1) show clearer bands (white) of temporal regularity in the phase-locked activity in the FFR at the fundamental period ($1/f_0$) and its multiples as compared with that of the English group (column 2). Similarly, the spectrograms of the Chinese group (column 3) show a markedly improved (white) spectral band at the $f_0$ and its harmonics as compared with the English group (column 4). No discernible differences are observed in response to T2tl, T2l or T2i between language groups.

board. Spectrograms reveal energy bands corresponding to $f_0$ and up to the fifth harmonic in the Chinese group in response to T2, but only up to the third or fourth in the English. In response to the other three stimuli, energy bands were more diffuse and less distinct regardless of language group.

### Pitch-tracking accuracy

An omnibus analysis of variance (ANOVA) on pitch tracking for curvilinear stimuli (T2, T2i) yielded a significant group × stimuli interaction effect [$F(1,18)=7.44$, $P=0.0138$]. In the case of T2 (Table 1), simple main effects revealed that pitch tracking was more accurate in the Chinese than in the English group [$F(1,18)=9.41$, $P=0.0066$], whereas group effects failed to reach significance in the case of T2i [$F(1,18)=0.69$, $P=0.4161$]. An omnibus ANOVA on pitch tracking for linear stimuli (T2l, T2tl) yielded no significant main or interaction effect.

**Table 1 Cross-correlation coefficient values of pitch tracking accuracy per language group and iterated rippled noise homologues of Mandarin Tone 2 and its variants**

| | Group | |
|---|---|---|
| Tone | Chinese | English |
| T2[a] | 0.91 (0.03) | 0.59 (0.06) |
| T2tl | 0.44 (0.10) | 0.45 (0.09) |
| T2l | 0.62 (0.12) | 0.44 (0.11) |
| T2i | 0.31 (0.08) | 0.40 (0.11) |

Values are expressed as mean and standard error (in parentheses).
T2, Mandarin Tone 2; T2i, inverted polynomial variant of T2; T2l, linear variant of T2; T2tl, trilinear variant of T2.
[a]Statistically significant difference between language groups in pitch-tracking accuracy.

Within each group, a direct comparison between T2 (curvilinear) and T2tl (trilinear) showed that pitch tracking of T2 was more accurate than T2tl in the Chinese group only [F(1,18)=19.66, $P$=0.0003; *c.f.* English, F(1,18)=1.62, $P$=0.2188].

### Pitch strength

FFR pitch strength is shown for six sections within each of the four IRN pitch patterns (Fig. 3). For curvilinear stimuli, an omnibus ANOVA on pitch strength in T2 yielded significant main effects of group [F(1,18)=19.79, $P$=0.0003] and section [F(5,90)=4.87, $P$=0.0005], but no group × section interaction. In T2, the Chinese group exhibited greater pitch strength than the English across sections. In T2i, an omnibus ANOVA yielded no significant main or interaction effects.

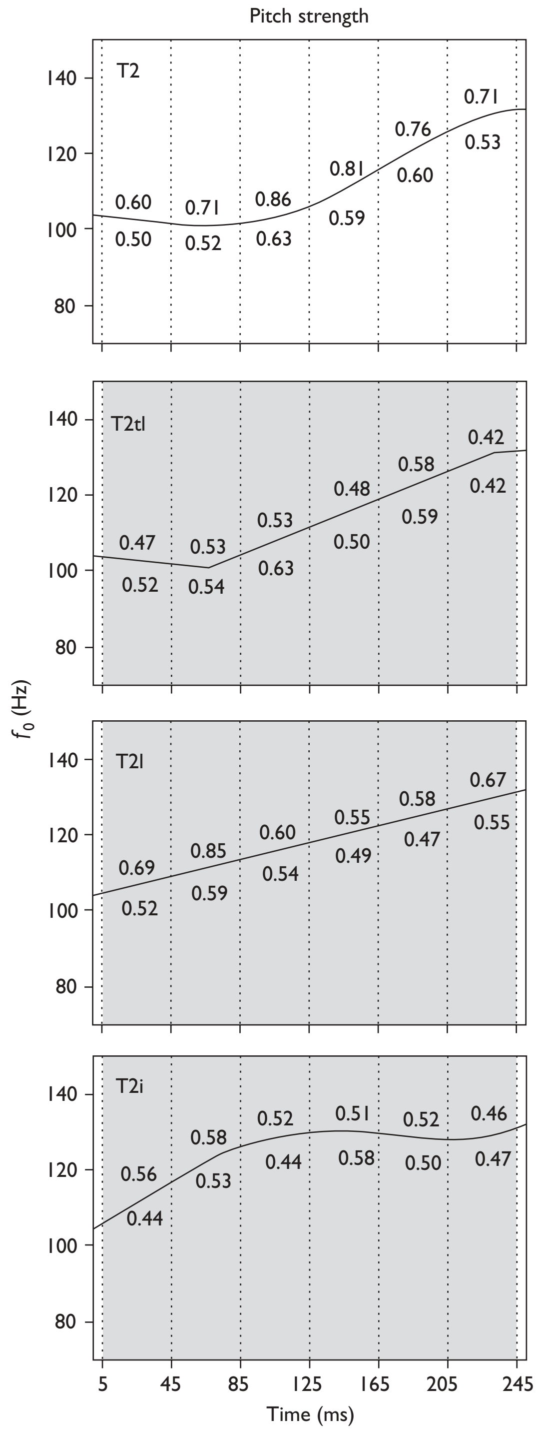
For linear stimuli, an omnibus ANOVA on pitch strength in T2l yielded a significant main effect of section [F(5,90)=3.23, $P$=0.0100]. Post-hoc Tukey–Kramer adjusted multiple comparisons ($\alpha$=0.05) showed that pitch strength of section 2 was greater than that of 4–5 across groups. Pitch strength in T2tl yielded a significant main effect of section [F(5,90)=3.99, $P$=0.0026]. Post-hoc multiple comparisons showed that pitch strength of sections 3 and 5 was greater than that of 6 across groups.

Within each group, a direct comparison of pitch strength of corresponding sections between T2 (curvilinear) and T2l (linear), the latter characterized by a fixed slope throughout its duration, yielded a significant section effect in the Chinese group only [F(11,107)=3.33, $P$=0.0006; *c.f.* English, F(11,107)=0.99, $P$=0.4577]. A priori contrasts ($\alpha$=0.05) indicated that pitch strength was greater in sections 3–5 of T2 as compared with corresponding sections of T2l.

## Discussion

The major finding of this cross-language study is that experience-dependent neural mechanisms for pitch representation at the brainstem level are sensitive to specific time-varying features of dynamic curvilinear pitch patterns that native speakers of a tone language are exposed to. The absence of language group effects in response to curvilinear (T2i) as well as linear (T2l, T2tl) variants of T2 emphasizes that language-dependent neuroplasticity at the level of the brainstem extends only to those pitch patterns that actually occur in the Mandarin tonal space. In a direct comparison between corresponding sections of T2 and T2l, pitch strength of the Chinese group, but not the English, differs especially in those sections exhibiting higher degrees of acceleration (*c.f.* Fig. 1: sections 3–5 vs. 1–2, 6). These findings extend our previous cross-language comparisons of FFRs elicited by curvilinear pitch patterns presented in a speech and

**Fig. 3**




Pitch strength of iterated rippled noise homologues of Mandarin Tone 2 and its variants within each of their six sections derived from the averaged frequency following response waveforms of Chinese and English participants (T2, row 1; T2tl, trilinear variant of T2, row 2; T2l, linear variant of T2, row 3; T2i, inverted polynomial variant of T2, row 4). Pitch strength of the Chinese group (value above solid line) is greater than that of the English group (value below solid line) for T2. Sections that yield significantly larger pitch strength for the Chinese group relative to the English group are unshaded; those that do not are shaded in grey. Vertical dotted lines demarcate six 40 ms sections: 5–45, 45–85, 85–125, 125–165, 165–205 and 205–245.

nonspeech contexts [10–12], and by linear pitch patterns presented in a speech context [17].

Language experience effects are not observed regardless of how close a 'linear' pitch pattern (T2tl) approximates a native curvilinear lexical tone (T2). Even 'curvilinear' pitch patterns (T2i) *per se* are insufficient to elicit language group effects. Language experience-dependent neuroplasticity occurs only when salient dimensions of pitch present in the auditory signal are part of the listener's experience and relevant to speech perception.

The degree of acceleration (and deceleration, *c.f.* [10]) of pitch trajectories seems to be a critical variable that influences pitch extraction in the rostral brainstem. We hypothesize that cross-language differences in the sustained phase-locked activity of the brainstem reflect an enhancement of selectivity to pitch-relevant periodicities that correspond to rapidly changing dynamic portions of the pitch contour. The role of the brainstem is to facilitate cortical level processing of pitch-relevant information by optimally capturing those features of the auditory signal that are of linguistic relevance.

To explain the brainstem mechanism underlying FFR pitch extraction and how language experience may alter this mechanism, we adopt the temporal correlation analysis model described by Langner [20,21]. Coincidence detection neurons in the inferior colliculus perform a correlation analysis on the delayed and undelayed temporal information from the cochlear nucleus to extract pitch-relevant periodicities that are spatially mapped onto a periodicity pitch axis. This encoding scheme is accomplished by neurons with different best modulation frequencies arranged in an orderly fashion orthogonal to the tonotopic frequency map. Its sensitivity can be enhanced by long-term experience, as reflected by smoother tracking of pitch contours and greater pitch strength, thereby sharpening the tuning characteristics of the best modulation frequency neurons along the pitch axis with particular sensitivity to linguistically relevant features.

Novel signal-processing algorithms have been proposed to enhance efficacy of cochlear implants (CI) for use with tone languages [22–24]. The FFR can faithfully preserve dynamic time-varying features critical for tonal languages, and can serve as a noninvasive neural index to evaluate different tonal CI signal processing strategies. A sectional analysis of the FFR suggests that CI algorithms be able to encode information at specific time-varying portions of auditory input, which are critical to neurophysiological representations of pitch. The relatively poor performance of current CI algorithms with respect to encoding of music and lexical tones may reflect a failure to optimally encode these dynamic pitch changes.

## Conclusion

Experience-dependent enhancement of pitch representation at the brainstem level is specific to pitch patterns that occur within the native listeners' experience. The absence of a language experience effect in response to the linear (T2l, T2tl) as well as nonnative curvilinear (T2i) pitch patterns is consistent with the notion that neural mechanisms underlying experience-dependent selectivity are local to the generators of the FFR in the human brainstem. Dynamic IRN stimuli permit us to investigate the degree of linguistic specificity of subcortical neural processing of pitch patterns minus a semantic confound.

## Acknowledgements

This study was supported by NIH R01 DC008549 (A.K.); NIH predoctoral traineeship (G.B.).

## References

1 Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci* 2007; **8**:393–402.
2 Poeppel D, Idsardi WJ, Van Wassenhove V. Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 2008; **363**:1071–1086.
3 Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 2008; **363**:1087–1104.
4 Johnson KL, Nicol TG, Kraus N. Brain stem response to speech: a biological marker of auditory processing. *Ear Hear* 2005; **26**:424–434.
5 Kraus N, Banai K. Auditory-processing malleability: focus on language and music. *Curr Dir Psychol Sci* 2007; **16**:105–110.
6 Kraus N, Nicol T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci* 2005; **28**:176–181.
7 Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci U S A* 2007; **104**:15894–15898.
8 Song JH, Skoe E, Wong PCM, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 2008; **20**:1892–1902.
9 Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 2007; **10**:420–422.
10 Krishnan A, Swaminathan J, Gandour JT. Experience dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* (in press). doi: 10.1162/jocn.2009.21077.
11 Krishnan A, Xu Y, Gandour JT, Cariani P. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 2005; **25**:161–168.
12 Swaminathan J, Krishnan A, Gandour JT. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 2008; **19**:1163–1167.
13 Krishnan A. Human frequency following response. In: Burkard RF, Don M, Eggermont JJ, editors. *Auditory evoked potentials: basic principles and clinical application*. Baltimore: Lippincott Williams & Wilkins; 2006. pp. 313–335.
14 Patterson RD, Handel S, Yost WA, Datta AJ. The relative strength of the tone and noise components in iterated ripple noise. *J Acoust Soc Am* 1996; **100**:3286–3294.
15 Yost WA, Patterson R, Sheft S. A time domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am* 1996; **99**:1066–1078.
16 Swaminathan J, Krishnan A, Gandour JT. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Trans Biomed Eng* 2008; **55**:281–287.
17 Xu Y, Krishnan A, Gandour JT. Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 2006; **17**:1601–1605.
18 Xu Y. Contextual tonal variations in Mandarin. *J Phon* 1997; **25**:61–83.

19 Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc Inst Phon Sci* 1993; **17**:97–110.
20 Langner G. Periodicity coding in the auditory system. *Hear Res* 1992; **60**:115–142.
21 Langner G. Topographic representation of periodicity information: the 2nd neural axis of the auditory system. In: Syka J, Merzenich M, editors. *Plasticity of the central auditory system and processing of complex acoustic signals*. New York: Plenum Press; 2004. pp. 21–26.
22 Lan N, Nie KB, Gao SK, Zeng FG. A novel speech-processing strategy incorporating tonal information for cochlear implants. *IEEE Trans Biomed Eng* 2004; **51**:752–760.
23 Luo X, Fu QJ. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *J Acoust Soc Am* 2004; **116**:3659–3667.
24 Nie K, Stickney G, Zeng FG. Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng* 2005; **52**:64–73.