

# Pitch encoding in speech and nonspeech contexts in the human auditory brainstem

Jayaganesh Swaminathan, Ananthanarayan Krishnan and Jackson T. Gandour

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

Correspondence to Professor Jackson T. Gandour, 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 3821; fax: +1 765 494 0771; e-mail: gandour@purdue.edu

Received 20 April 2008; accepted 12 May 2008

Frequency-following responses were recorded from Chinese and English participants at the level of the brainstem in response to four Mandarin tonal contours presented in a speech and nonspeech context. Pitch strength analysis of these preattentive brainstem responses showed that the Chinese group exhibited stronger pitch representation than the English group regardless of context. Moreover, the Chinese group exhibited relatively more robust pitch representation of rapidly changing pitch segments. These

findings support the view that at early preattentive stages of subcortical processing, neural mechanisms underlying pitch representation are shaped by particular features of the auditory stream rather than speech *per se*. These findings have implications for optimizing signal-processing strategies for cochlear implant design for speakers of tonal languages. *NeuroReport* 19:1163–1167 © 2008 Wolters Kluwer Health | Lippincott Williams & Wilkins.

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

## Introduction

Languages that exploit variations in pitch-to-signal meaning differences in monosyllabic words are called tone languages. By using scalp-recorded human frequency-following responses (FFR), it has been shown that preattentive stages of pitch encoding of Mandarin tones are sensitive to language experience at the level of the human auditory brainstem [1]. Pitch information is preserved in the phase-locked neural activity generating the FFR not only for steady-state complex tones [2], but also for time-varying pitch contours of Mandarin speech [3]. Thus, the FFR provides a noninvasive electrophysiological measure of neural phase locking and serves as an optimal window to view neural processing of pitch at the level of the auditory brainstem. It also serves as a useful tool to probe in to questions related to experience-dependent plasticity and preattentive lower level of sensory processing on pitch.

To generate auditory stimuli that preserve the perception of pitch, minus waveform periodicity or highly modulated stimulus envelopes, we use iterated rippled noise (IRN). An IRN stimulus is generated using a broadband noise that is delayed and added to itself repeatedly. The perceived pitch corresponds with the reciprocal of the delay, and the pitch salience increases with the number of iterations of the delay-and-add process [4,5]. The IRN algorithm has been generalized to allow multiple time-dependent delays over a range of iteration steps, making it possible for humans to detect pitch changes in 'dynamic' IRN by humans [6], and further modified to handle curvilinear pitch contours ecologically representative of natural speech [7].

At an early preattentive 'subcortical' stage of processing, FFRs elicited in response to Mandarin tones reveal smoother

pitch tracking in native versus nonnative listeners, no matter the context, speech or nonspeech [1,8]. By measuring 'pitch strength', peak of autocorrelation function, we are able to focus on individual sections of pitch contours. The primary aim of this cross-language study was to determine whether the brainstem mechanisms responsible for extracting pitch information are susceptible to stimulus degradation. Specifically, do FFRs induce more robust phase locking to speech than nonspeech stimuli with degraded periodicity (e.g. IRN)? Another aim was to determine whether experience-dependent neural mechanisms for pitch representation in the brainstem are sensitive to specific time-varying features of pitch contours that native speakers of a tone language are familiar with regardless of context.

## Methods

### Participants

Fourteen adult native speakers of Mandarin Chinese and 13 native speakers of American English participated in the Mandarin 'speech' experiment [1]. Separate groups of 12 adult native speakers of Mandarin and 12 adult monolingual native speakers of English participated in the Mandarin 'nonspeech' experiment. Participants' ages ranged from 21 to 30 years. All Chinese participants were born and raised in Mainland China. They gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

### Stimuli

In the speech experiment, a set of Mandarin monosyllables was chosen to contrast the four lexical tones: /y<sup>i</sup>/

'clothing', /y<sup>i</sup>/ 'aunt', /y<sup>i</sup><sup>3</sup>/ 'chair', /y<sup>i</sup><sup>4</sup>/ 'easy'.  $F_0$  contours were modeled after natural productions of citation forms. In the nonspeech experiment, time-varying IRN stimuli were created with the same  $f_0$  contours at a high-iteration step ( $n=32$ ) using procedures described in the study by Swaminathan *et al.* [7]. Stimulus duration was 250 ms including a 10-ms cosine squared ramp used to eliminate both spectral splatter and artifactual onset responses.

### Data acquisition

The data acquisition procedures are as described in the study by Krishnan *et al.* [1]. FFRs were recorded from each participant in response to monaural stimulation of the right ear. In the speech experiment, these evoked responses were recorded differentially between scalp electrodes placed on the midline of the forehead at the hairline and the seventh cervical vertebra (C7). Another electrode placed on the mid-forehead (Fpz) served as the common ground. In the nonspeech experiment, the FFRs were recorded differentially between scalp electrodes placed on the midline of the forehead at the hairline and the ipsilateral mastoid. Another electrode placed on the contralateral mastoid served as the common ground. These two derivations yielded essentially the same responses.

### Data analysis

#### Pitch strength of tonal sections

To compute the pitch strength of the FFR responses to speech and nonspeech stimuli, 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 speech and nonspeech stimuli and their FFR responses. The first author identified visually the location of the autocorrelation peak per 40-ms frame from the input IRN stimuli. This location was then used to guide a visual search for the corresponding peak in the FFR response. 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. This response peak was taken to be an estimate of pitch strength per time frame.

### Results

FFR pitch strength, as measured by the average magnitude of the normalized autocorrelation peak per language group (Fig. 1) and context (Fig. 2), is shown for six sections within each of the four IRN homologs of Mandarin tones. Across the four tones (Fig. 1), pitch strength in the speech context is observed to be greater than in the nonspeech context in 83 and 92% of sections for the Chinese and the English groups, respectively. Seventy-five percent of overlap between groups in those tonal sections (unshaded) in which we observe a context effect is seen. Across the four tones (Fig. 2), pitch strength of the Chinese group is observed to be significantly greater than the English group in nearly twice as many tonal sections (unshaded) in the nonspeech (15) as in the speech (seven) context.

For each tone separately, results from an omnibus three-way (group  $\times$  context  $\times$  section) analysis of variance performed on pitch strength revealed significant ( $P < 0.0001$ ) main effects of group [T1:  $F(1,240)=24.96$ ; T2=53.10;

T3=30.57; T4=50.89], context [T1:  $F(1,240)=24.23$ ; T2=55.58; T3=83.37; T4=14.62], and section [T1:  $F(5,240)=8.02$ ; T2=10.95; T3=10.69; T4=5.61]. The context  $\times$  section interaction was significant for T1 [ $F(1,5)=5.45$ ,  $P < 0.0001$ ]. For T3, all two-way interactions were significant: context  $\times$  section [ $F(1,5)=7.79$ ,  $P < 0.0001$ ]; context  $\times$  group [ $F(1,1)=5.61$ ,  $P < 0.0186$ ]; group  $\times$  section [ $F(1,5)=2.39$ ,  $P < 0.0385$ ]. No other two-way or three-way interaction effects reached significance.

For each tone and group (Fig. 1), a two-way analysis of variance of pitch strength revealed significant main effects of section and context across all four tones ( $P < 0.05$ ). The context  $\times$  section interaction was significant in the majority of cases except for T2 in the Chinese group and T2 and T3 in the English group. Regardless of language group, Tukey-adjusted comparisons indicated that in 16 out of 24 tonal sections, pitch strength in the speech context was greater than in the nonspeech context (Fig. 1, unshaded,  $P < 0.05$ ). On an average, the pitch strength in the speech context for the Chinese and English group, respectively, was 1.25 and 1.45 times greater than that in the nonspeech context. Of those tonal sections in which the reverse pattern occurred (i.e. nonspeech more than speech), all (four) occurred at the beginning of the IRN stimulus.

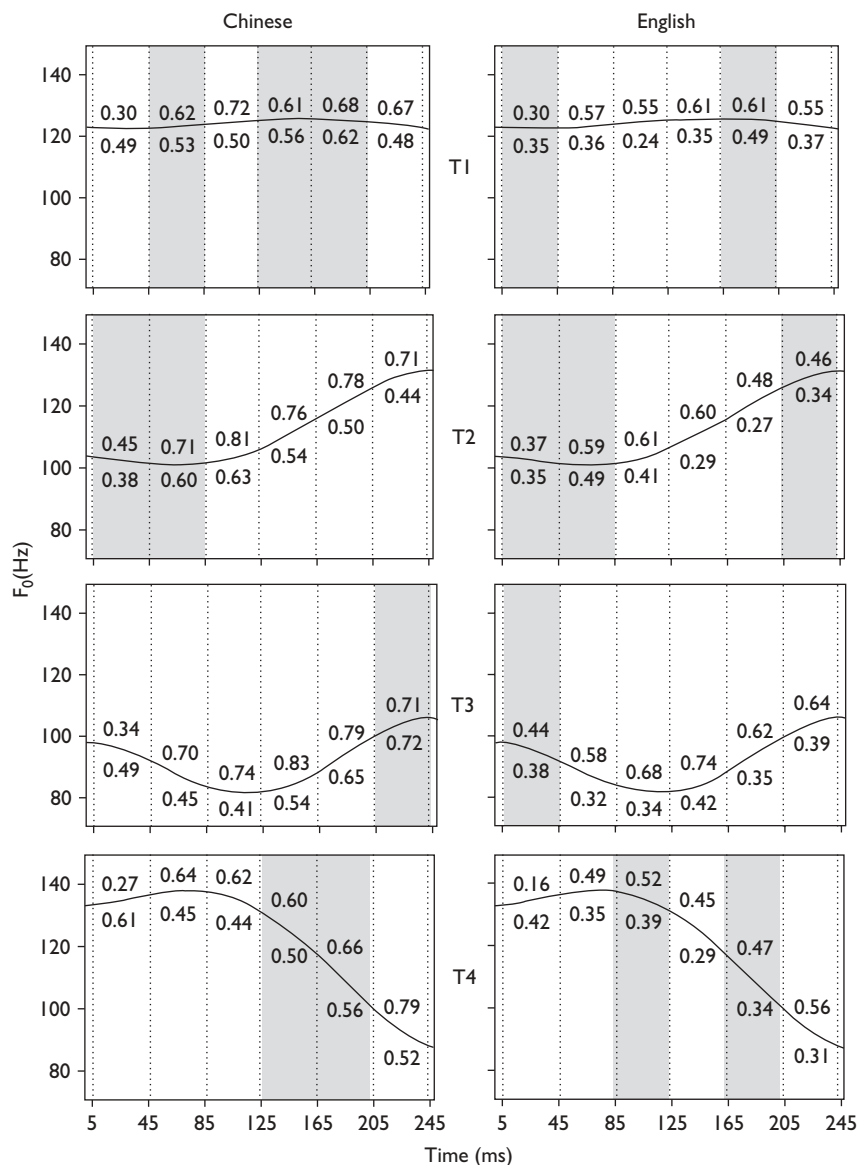
For each tone and context (Fig. 2), a two-way analysis of variance of pitch strength revealed significant main effects of section and group across all four tones ( $P < 0.05$ ). No two-way or three-way interaction effects reached significance. Regardless of context, Tukey-adjusted comparisons indicated that in seven out of 24 and 15 out of 24 tonal sections in the speech and nonspeech contexts, respectively, pitch strength in the Chinese group was greater than in the English group (Fig. 2, right panels, unshaded,  $P < 0.05$ ).

Table 1 presents the acceleration values of the six sections of each of the four IRN homologs of Mandarin tones. Pooling across tones, a positive correlation coefficient was observed between the pitch strength ratios of the two language groups and acceleration (absolute) values of the Mandarin pitch contours per section in both speech ( $r=0.37$ ,  $P=0.0270$ ) and nonspeech ( $r=0.45$ ,  $P=0.075$ ) contexts.

### Discussion

The major finding of this cross-language study is that independent of the speech–nonspeech context, experience-dependent neural mechanisms for pitch representation at the brainstem level are sensitive to specific time-varying features of pitch patterns that native speakers of a tone language are exposed to. We infer that 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. Dynamic IRN stimuli permit us to investigate neural mechanisms underlying pitch patterns representative of those that occur in natural speech without a semantic confound.

We also observed greater pitch strength for speech compared with nonspeech stimuli for both English and Chinese listeners. The weaker pitch strength for the nonspeech stimuli is to be expected given the relatively less robust temporal periodicity in the stimulus waveform. Nevertheless, our data indicate that dynamic IRN stimuli do preserve fine-grained measures of pitch representation at the level of the brainstem, thus giving us a window on neural representation of pitch in degraded conditions.

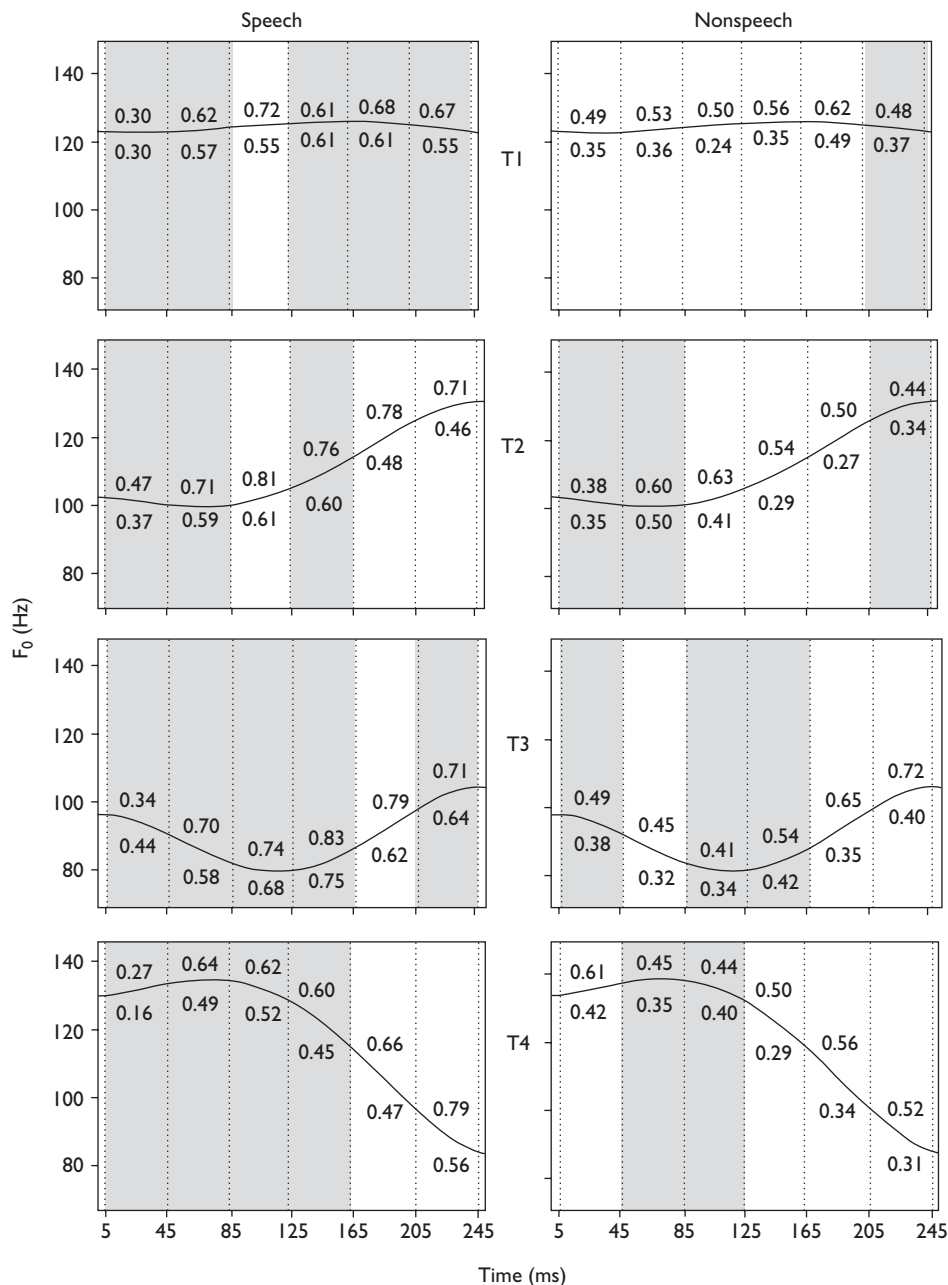


**Fig. 1** Pitch strength of tonal sections derived from the frequency-following response waveforms of Chinese (left) and English participants (right) in response to speech and nonspeech stimuli. The four Mandarin tonal categories are represented by T1, T2, T3, and T4. Consistent across both language groups, in the majority of sections, pitch strength derived in response to speech stimuli (value above the solid line) is greater than response to nonspeech stimuli (value below the solid line). Sections that yield significantly larger pitch strength for the speech stimuli relative to nonspeech stimuli are unshaded; those that are not shaded in gray. Vertical dotted lines demarcate six 40-ms sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245.

Regardless of context, pitch strength of the Chinese group is greater than that of the English (Fig. 2). Group differences in pitch strength are, however, not uniform throughout the duration of FFR responses to either speech or their IRN homologs. It is observed that in some tonal sections that have rapid changes (e.g. T4, S4; speech), the two language groups do not differ in pitch strength. Conversely, in other tonal sections that are relatively smooth (e.g. T1, S3; speech), pitch strength differs between the two groups. Nonetheless, we infer that neural mechanisms in the brainstem are not responding to lexical tones *per se*, but rather to specific time-varying acoustic properties of the input stimuli. The degree of acceleration and deceleration of the pitch trajectories

seems to be a critical variable that influences pitch extraction in the rostral brainstem. Pitch strength differs as a function of language experience especially in those tonal sections exhibiting higher degrees of acceleration (e.g. Table 1: T3, S5) and deceleration (e.g. Table 1: T4, S5). 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 with rapidly changing dynamic segments of the pitch contour.

Novel signal processing algorithms have recently been proposed to enhance efficacy of cochlear implants (CI) for use with tone languages [9–11]. Although they have



**Fig. 2** Pitch strength of tonal sections derived from the frequency-following response waveforms in response to speech (left) and nonspeech (right) stimuli for the two language groups. The four Mandarin tonal categories are represented by T1, T2, T3, and T4. Consistent across speech and nonspeech stimuli, the pitch strength of the Chinese group (value above the solid line) is greater than that of the English group (value below the solid line). Sections that yield significantly larger pitch strength for the Chinese group relative to English (unshaded) are those sections that exhibit larger acceleration or deceleration values (cf. Table 1). Vertical dotted lines demarcate six 40-ms sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245.

been tested perceptually with normal hearing and deaf CI patients, there are, as of yet, no physiological data to show an improvement in neural representation of time-varying features [12–14]. 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 representa-

tions of pitch. Such a neural index would facilitate development and testing of optimal CI algorithms that preserve critical time-varying portions of the pitch.

In conclusion, our findings demonstrate that experience-dependent neural mechanisms for pitch representation at the brainstem level are not speech specific but instead are sensitive to ‘specific dimensions’ of pitch contours that native speakers of a tone language are familiar with. We infer that the role of the brainstem is to facilitate cortical level processing of pitch relevant information by ‘optimally’

**Table 1** Acceleration values of the six sections from each of the four IRN homologs of Mandarin tones

Tone	Section					
	S1	S2	S3	S4	S5	S6
T1	-0.0002	0.0013	0.0014	0.0005	-0.0008	-0.0022
T2	-0.0023	0.0001	0.0049	0.0095	0.0108	0.0058
T3	-0.0059	-0.0086	-0.0023	0.0066	0.0118	0.0068
T4	0.0034	0.0011	-0.0063	-0.0143	-0.0181	-0.0131

S1, S2, S3, S4, S5, and S6 represent the six 40-ms sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245. T1, T2, T3, and T4 stand for the four Mandarin tones. Values represent the degree of acceleration/deceleration, defined as rate of change in pitch, within each section. For a 40-ms time frame, acceleration was computed as the difference in pitch value at offset and onset divided by the duration of the frame. Positive and negative signs represent rising and falling  $f_0$  trajectories, respectively. IRN, iterated rippled noise.

capturing those dimensions of the auditory signal that are of linguistic relevance.

### Acknowledgement

Sources of support: NIH R01 DC008549-01 (A.K.); College of Liberal Arts (A.K., J.G.).

### References

- 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.
- Greenberg S, Marsh JT, Brown WS, Smith JC. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear Res* 1987; **25**:91–114.
- Krishnan A, Xu Y, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004; **189**:1–12.
- 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.
- Yost WA. Pitch strength of iterated rippled noise. *J Acoust Soc Am* 1996; **100**:3329–3335.
- Denham S. Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model. *Biosystems* 2005; **79**:199–206.
- 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.
- 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).
- 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.
- Luo X, Fu QJ. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *J Acoust Soc Am* 2004; **116**:3659–3667.
- Nie K, Stickney G, Zeng FG. Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng* 2005; **52**:64–73.
- Fu QJ, Hsu CJ, Horng MJ. Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant users. *Ear Hear* 2004; **25**:501–508.
- Hsu CJ, Horng MJ, Fu QJ. Effects of the number of active electrodes on tone and speech perception by Nucleus 22 cochlear implant users with SPEAK strategy. *Adv Otorhinolaryngol* 2000; **57**:257–259.
- Liu SY, Huang TS, Follent M. The field trial of the SPEAK versus MPEAK speech coding strategies in Mandarin Chinese. *Adv Otorhinolaryngol* 1997; **52**:113–116.