This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Short Communication

Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech

Ananthanarayan Krishnan a, Jackson T. Gandour a, *, Christopher J. Smalt b, Gavin M. Bidelman a

a Department of Speech Language Hearing Sciences, Purdue University, USA
b School of Electrical and Computer Engineering, Purdue University, USA

A R T I C L E   I N F O

Article history:
Accepted 12 May 2010
Available online 8 June 2010

Keywords:
Auditory
Human
Brainstem
Pitch
Language
Frequency-following response (FFR)
Click trains
Mandarin Chinese
Experience-dependent plasticity
Speech perception

A B S T R A C T

Experience-dependent enhancement of neural encoding of pitch in the auditory brainstem has been observed for only specific portions of native pitch contours exhibiting high rates of pitch acceleration, irrespective of speech or nonspeech contexts. This experiment allows us to determine whether this language-dependent advantage transfers to acceleration rates that extend beyond the pitch range of natural speech. Brainstem frequency-following responses (FFRs) were recorded from Chinese and English participants in response to four, 250-ms dynamic click-train stimuli with different rates of pitch acceleration. The maximum pitch acceleration rates in a given stimulus ranged from low (0.3 Hz/ms; Mandarin Tone 2) to high (2.7 Hz/ms; 2 octaves). Pitch strength measurements were computed from the FFRs using autocorrelation algorithms with an analysis window centered at the point of maximum pitch acceleration in each stimulus. Between-group comparisons of pitch strength revealed that Chinese exhibit more robust pitch representation than English across all four acceleration rates. Regardless of language group, pitch strength was greater in response to acceleration rates within or proximal to natural speech relative to those beyond its range. Though both groups showed decreasing pitch strength with increasing acceleration rates, pitch representations of the Chinese group were more resistant to degradation. FFR spectral data were complementary across acceleration rates. These findings demonstrate that perceptually salient pitch cues associated with lexical tone influence brainstem pitch extraction not only in the speech domain, but also in auditory signals that clearly fall outside the range of dynamic pitch that a native listener is exposed to.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

There is an emerging literature to support the notion that the neural representation of pitch may be influenced by one’s experience with language (or music) at subcortical as well as cortical levels of processing (for reviews, see Johnson, Nicol, & Kraus, 2005; Kraus & Banai, 2007; Kraus & Nicol, 2005; Krishnan & Gandour, 2009; Patel & Iversen, 2007; Tzounopoulos & Kraus, 2009; Zatorre & Gandour, 2008). Pitch provides an excellent window for studying language-dependent effects on subcortical processing as it is one of languages’ most important information-bearing components. Tone languages are especially advantageous for investigating the linguistic use of pitch because variations in pitch patterns at the syllable level may be lexically significant (Yip, 2003). Mandarin Chinese, for example, has four lexical tones: ma1 ‘mother’, ma2 ‘hemp’, ma3 ‘horse’, ma4 ‘scold’ (Howie, 1976).

As a window into pitch processing in the brainstem, we measure electrophysiological activity using the human frequency-following response (FFR). This response reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem (see Krishnan (2006) for review). The FFR is characterized by a periodic waveform that follows the individual cycles of the stimulus waveform. FFRs can be elicited by a variety of stimuli that carry periodicity information, including low frequency tone bursts, complex tones, speech sounds, sinusoidal amplitude modulated sounds, frequency modulated sounds, click trains, and iterated rippled noise (IRN). Experimental evidence overwhelmingly points to the inferior colliculus (IC) as the source of the FFR generator.

As reflected by FFRs, comparisons between native speakers of tone (Mandarin) and non-tone (English) languages show that native experience with lexical tones enhances pitch encoding at the level of the brainstem irrespective of speech or nonspeech context (Krishnan, Gandour, Bidelman, & Swaminathan, 2009; Krishnan, Xu, Gandour, & Cariati, 2005; Swaminathan, Krishnan, & Gandour, 2008). Language-dependent pitch encoding mechanisms in the brainstem are especially sensitive to the curvilinear shape of pitch contours that are exemplary of those that occur in natural speech.
We fail to observe any language-dependent effects no matter how close a linearly accelerating or decelerating pitch pattern approximates a native lexical tone (Krishnan, Gandour, et al., 2009; Xu, Krishnan, & Gandour, 2006). Curvilinearity itself, though necessary, is insufficient to enhance pitch extraction of the auditory signal at the level of the brainstem. A nonnative curvilinear pitch pattern similarly fails to elicit a language-dependent effect (Krishnan, Gandour, et al., 2009). Using synthetic speech stimuli representative of Mandarin and Thai lexical tones, listeners from both tone languages are shown to be able to transfer their abilities in pitch encoding across languages (Krishnan, Gandour, & Bidelman, 2010b). Thus, brainstem neurons appear to be differentially sensitive to changes in specific dimensions of pitch without regard to their language identity as long as they occur in a language with a comparable phonological system. These findings collectively suggest that language-dependent neuroplasticity in the human brainstem occurs when dimensions of pitch in the auditory signal are part of the listener’s experience and relevant to speech perception. By analyzing pitch strength of individual sections of lexical tones, one of our major discoveries is that the degree of pitch acceleration or deceleration (i.e., rate of pitch change) is a critical dimension that influences pitch extraction in the brainstem. Chinese listeners exhibit more robust pitch representation than English primarily in those sections of lexical tones containing rapidly changing pitch regardless of the context, speech or nonspeech (Krishnan, Swaminathan, & Gandour, 2009). This heightened sensitivity to sections characterized by rapid changes in pitch is maintained even in severely degraded stimuli (Krishnan, Gandour, & Bidelman, 2010a). Thus, experience-dependent brainstem mechanisms for pitch are especially sensitive to those dimensions of pitch contours that provide cues of high perceptual saliency in degraded pitch stimuli.

2. Results

2.1. Temporal and spectral properties of whole stimuli

Autocorrelograms (left panels) and narrow-band spectrograms (right panels) derived from the grand averaged FFR waveforms in response to click-train f0 contours representative of Tone 2 (row 1), two intermediate rates of higher acceleration (rows 2–3), and an extraordinarily high acceleration rate (row 4) are shown in Fig. 1 for the Chinese and English groups (cf. Fig. 3). In the Chinese group, autocorrelograms show clear dark bands of phase-locked activity at f0 and its multiples in response to an f0 contour characterized by a linguistically-relevant acceleration rate (row 1), but less distinct and increasingly more diffuse bands in response to f0 contours of greater acceleration (rows 2–4). In the English group, the bands are less distinct and more diffuse at all acceleration levels. At the highest acceleration rate, A4, the English group’s correlogram appears to have no bands whatsoever in the region of interest (170 ms), unlike the Chinese group.

Consistent with the autocorrelograms, spectrograms in Fig. 1 shows more robust representation of the first five harmonics, particularly for the higher acceleration rates for the Chinese group. Stimulus-to-response spectral correlation coefficients within the region of maximum acceleration are reported for each of four acceleration rates per language group (Table 1). Results from an omnibus two-way (group × acceleration) ANOVA on spectral correlation coefficients yielded significant main effects of group (F1,18 = 20.57, p = 0.0003) and acceleration (F3,54 = 4.87, p < 0.0045). The two-way interaction failed to reach significance (F3,54 = 1.07, p = 0.3699), meaning that the Chinese group response spectrum was more correlated with the stimulus spectrum as compared to the English across all acceleration rates, suggesting that the representation of pitch-relevant harmonics were more robust in the Chinese group for all stimuli.

2.2. Pitch strength of region of interest

FFR pitch strength of the region of maximum acceleration within the click-train stimuli is shown for each of four acceleration

...
rates per language group (Fig. 2) (Supplementary material: table; pitchstrength,M+SE.doc). Results from an omnibus two-way (group x acceleration) ANOVA on pitch strength yielded significant main effects of group ($F_{1,18} = 19.11$, $p = 0.0004$) and acceleration ($F_{3,54} = 27.66$, $p < 0.0001$). The two-way interaction failed to reach significance ($F_{3,54} = 1.53$, $p = 0.2184$), meaning that pitch strength was greater in the native Chinese group as compared to the English across the board. Within both groups, post hoc Tukey multiple comparisons ($\alpha = 0.05$) revealed that pitch strength was greater at lower (A1, A2) than higher (A3, A4) acceleration rates. In the English group, pitch strength was also greater in A3 as compared to A4.

### 3. Discussion

In this crosslanguage study, native speakers of a tone language, relative to those of a non-tone language, are observed to exhibit more robust pitch representation within tonal sections exhibiting maximal rates of pitch change regardless of their behavioral relevance. As measured by pitch strength, not only do the Chinese have an advantage over English listeners in response to an exemplar of Mandarin Tone 2, but also to scaled variants of T2 with increasingly higher maximum acceleration rates that fall proximal to or outside the boundary of natural speech. This finding demonstrates that neuroplasticity for pitch processing in the brainstem is not limited to the domain in which the pitch contours are behaviorally relevant (cf. Bidelman et al., in press). Irrespective of language experience, there are substantial drops in sensitivity to those acceleration rates that extend well beyond the normal voice range (A1, A2 > A3, A4). Yet Chinese are still three times more sensitive than English in
response to the pitch contour with the highest rate of pitch acceleration (A4).

3.1. Neural mechanisms underlying pitch encoding in the brainstem are differentially sensitive to increasing rates of pitch acceleration as a function of language experience

Our view is that crosslanguage differences in pitch representation reflect experience-dependent sensitivity of the neural mechanism underlying pitch encoding in native speakers of tone languages (Krishnan & Gandour, 2009). We have proposed that coincidence detection neurons in the inferior colliculus perform a correlation analysis (Langner, 1992, 2004) to extract pitch-relevant periodicities on the delayed and un-delayed temporal information arriving from the cochlear nucleus. This pitch information is then spatially mapped onto a periodicity pitch axis. We believe that long-term experience sharpens the tuning characteristics of the best modulation frequency neurons along the pitch axis with particular sensitivity to linguistically-relevant dynamic segments. This sharpening is likely mediated by local excitatory and inhibitory interactions that are known to play an important role in signal selection at the level of the brainstem (Ananthanarayan & Gerken, 1983, 1987). Such interaction may take the form of an active facilitation/disinhibition of the pitch intervals corresponding to the dynamic segments and inhibition of other pitch periods. Neuronal modulatory inputs to the corticocollicular loop could influence the balance between excitation and inhibition (Xiong, Zhang, & Yan, 2009).

We observe herein a degradation of pitch representation with increasing pitch acceleration across language groups. This suggests a disruption in the correlation analysis to extract pitch-relevant periodicities. One plausible explanation for the disruption of this correlation analysis is that the coherence of the phase-locked neural activity representing pitch-relevant periodicities decreases with increasing acceleration of the fundamental frequency. This reduced coherence is due to temporal de-synchronization of the inputs to the pitch extraction mechanism. The fact that we also observe more robust pitch representation in the Chinese listeners even at higher acceleration rates suggests that their pitch extraction mechanism is less susceptible to de-synchronization. Both enhanced sensitivity and decreased susceptibility to degradation of pitch encoding for faster pitch acceleration rates in the Chinese listeners suggests an expansion of the dynamic range for this dimension of pitch. We argue that their heightened sensitivity to rate-of-change of frequency, a direct result of their language experience, is a crucial factor distinguishing tone language from non-tone language speakers/listeners.

Another finding is that multiple harmonics are better represented in Chinese than English listeners even at higher acceleration rates. This finding complements our data on voice pitch representation (Krishnan et al., 2005). In the Chinese group, stronger pitch and more accurate pitch tracking co-occur with relatively stronger representation of pitch-relevant harmonics. Just the opposite is the case for the English group. Moreover, psychoacoustic and physiological data indicate that complex stimuli produce stronger and more accurate pitch percepts when spectral components are prominent in the dominance region for pitch (2nd to about the 5th harmonic) (Cariani & Delgutte, 1996; Schwartz & Purves, 2004).
3.2. Experience-dependent neuroplasticity in the brainstem is not circumscribed to the language domain

Background noise, competing sounds, or in this study, higher rates of pitch acceleration, represent a significant challenge for neural mechanisms encoding behaviorally relevant acoustic features of the target sound. The extent to which a particular dimension of pitch is resistant to such degraded listening conditions may serve as an index of its perceptual saliency. Regarding pitch acceleration, differential rates throughout the duration of a lexical tone are what determine its shape or contour. It is this pitch dimension that has been shown to be an important variable in separating tone from non-tone language speakers, as evidenced from perceptual judgments (Gandour, 1983), cortical evoked potentials (Chandrasekaran, Gandour, & Krishnan, 2007), and, of course, subcortical responses in the brainstem (Krishnan, Swaminathan, et al., 2009). Despite their lack of pitch experience with these higher acceleration rates, Chinese listeners, as compared to English, are sensitive to linguistic prominence (Xu, 1997) using a four-step stimulus continuum ranging from lower to higher contour of Mandarin Tone 2 as measured by the Edinburgh Handedness inventory (Oldfield, 1971). All participants exhibited normal hearing sensitivity (better than 15 dB HL in both ears) at octave frequencies from 500 to 4000 Hz. Native speakers of Mandarin were born and raised in mainland China and none had received formal instruction in music. A third of Mandarin listeners was left-handed (Oldfield, 1971), and 13% indicated some hearing loss in either ear.

4. Methods

4.1. Participants

Ten adult native speakers of Mandarin Chinese (6 male, 4 female), hereafter referred to as Chinese (C), and 10 adult monolingual native speakers of English without musical training (5 male, 5 female), referred to as English (E), participated in the FFR experiment. The three groups were closely matched in age (Chinese: M = 23.8, SD = 1.8; English: M = 24.7, SD = 3.2), years of formal education (Chinese: M = 17.2, SD = 1.6; English: M = 18.1, SD = 3.6). All were strongly right handed (M = 87%) as measured by the Edinburgh Handedness inventory (Oldfield, 1971). All participants exhibited normal hearing sensitivity (better than 15 dB HL in both ears) at octave frequencies from 500 to 4000 Hz. Native speakers of Mandarin were born and raised in mainland China and none had received formal instruction in English before the age of 9 (M = 11.7, SD = 1.1). The English group had no prior experience learning a tonal language. Each participant completed a music history questionnaire (Wong & Perrachione, 2007). All participants had no more than 3 years of formal music training (M = 0.7, SD = 0.8) on any combination of instruments and none had any training within the past 5 years. All participants were paid and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

4.2. Stimuli

A four-step stimulus continuum ranging from lower to higher rates of pitch acceleration was utilized to evoke brainstem responses (Fig. 3, top panel). One end of the pitch acceleration continuum was represented by the f0 contour of Mandarin Tone 2 as produced in citation form by a male speaker (Xu, 1997) using a four-order polynomial (A1: \( A = 31.2 \text{Hz}/0.38 \text{octaves} \) (Swaminathan, Krishnan, Gandour, & Xu, 2008). \( A \) represents the difference in \( f_0 \) between turning point (65 ms), defined as the minimum \( f_0 \) along the duration of the pitch contour, and offset (250 ms) expressed in Hertz and octaves. At the other end of the continuum is a scaled version of the Tone 2 f0 contour that clearly extends beyond the limits of the normal voice range (A4: \( A = 299.2 \text{Hz}/1.99 \text{octaves} \)). Of the two remaining \( f_0 \) contours with intermediate rates of acceleration, one represents a pitch pattern that is of marginal relevance linguistically (A2: \( A = 74.2 \text{Hz}/0.74 \text{octaves} \)); the other, a pitch pattern that does not occur within a normal voice range (A3: \( A = 149.2 \text{Hz}/1.48 \text{octaves} \)). This continuum was derived from a study on the maximum speed at which a speaker can voluntarily change pitch (Xu & Sun, 2002). The maximum velocity in a rising direction was reported to be 61.3 semitones per second (st/s). In this study, we measured velocity from turning point to tonal offset, an excursion time of 185 ms. As compared to 61.3 st/s, A1 (25.4 st/s) falls well within the physiological limits of speed of rising pitch changes; A2 (51.94 st/s) similarly falls under the maximum, but is beginning to approach the maximum limit for changes in rising pitch; both A3 (85.44 st/s) and A4 (129.5 st/s) go far beyond the limit for changes in rising pitch.

The stimuli were generated using click trains whose inter-click intervals were changed according to \( 1/f_0 \) for each polynomial (Supplementary material: audio files, mp3; click-train_stimuli.doc). All stimuli were low pass filtered at 3000 Hz, matched in RMS amplitude, and fixed in duration to 250 ms including a 10 ms rise/fall time (cos2 ramps).

The size of the analysis window for each stimulus was fixed at five pitch periods, instead of a constant time interval, resulting in a varying window length demarcated by width of the rectangles (top panel). This process was necessary to ensure the same temporal resolution for the autocorrelation spectral analysis for each stimulus across the pitch acceleration continuum. The center of each window (at 177 ms) corresponded to the peak of the pitch acceleration curve (Fig. 3, bottom panel). Note that the pitch acceleration curves differed across stimuli throughout their duration except for the turning point (65 ms).

4.3. Data acquisition

FFRs recording protocol and data analysis were similar to those described in previous publications from our laboratory (Bidelman et al., in press; Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009). Participants reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize movement artifacts. FFRs were recorded from each participant in response to monaural stimulation of the right ear at 80 dB SPL at a repetition rate of 2.76/s. The presentation order of the stimuli was randomized both within and across participants. Control of the experimental protocol was accomplished by a signal generation and data acquisition system (Tucker–Davis Technologies, System III). The stimulus files were routed through a digital to analog module and presented through a magnetically shielded insert earphone (Etymotic, ER-3A).

FFRs were recorded differentially between a non-inverting (positive) electrode placed on the midline of the forehead at the hairline (Fz) and inverting (reference) electrodes placed on (i) the right mastoid (A2); (ii) the left mastoid (A1); and (iii) the 7th cervical vertebra (C7). Another electrode placed on the mid-front (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 a response with a higher signal-to-noise ratio (Krishnan, Gandour, et al., 2009). All inter-electrode impedances were maintained below 1 kΩ. The EEG inputs were amplified by 200,000 and band-pass filtered from 75 to 1500 Hz (6 dB/octave roll-off, RC, response characteristics). Each response waveform represented the average of 3000 stimulus presentations over a 280 ms analysis window using a sampling rate of 24,414 kHz. The experimental protocol took about 100 min to complete.
4.4. Data analysis

Short-term autocorrelation functions (ACF) and running autocorrelations were computed for the FFRs to index variation in FFR periodicities over the duration of the response to time-varying, click-train stimuli. The autocorrelogram (ACG) represents the short-term autocorrelation function of windowed frames of a compound signal, i.e., ACG$(t, \tau) = X(t) \times X(t - \tau)$ for each time $t$ and time lag $\tau$. It is a three dimensional plot quantifying the variations in periodicity and pitch strength as a function of time. The horizontal axis represents the time at which single ACF “slices” are computed while the vertical axis represents their corresponding time lags, i.e., pitch periods. The intensity of each point in the image represents the instantaneous ACF magnitude computed at a given time within the response.

Neural pitch strength was extracted from the FFRs using autocorrelation analysis (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009; Swaminathan, Gandour, Gandour, et al., 2008) applied to a time window centered around the region of maximum acceleration (Fig. 3, bottom panel). This analysis yielded estimates of both pitch period (time lag associated with the autocorrelation maximum) and pitch strength (magnitude of the normalized autocorrelation expressed as a value between 0 and 1 where, 0 represents an absence of periodicity and 1 represents maximal periodicity). It should be noted here that the time window used for the autocorrelation analysis varied in duration across the stimuli in order to include exactly five pitch periods in the FFR response – a necessary requirement to maintain constant temporal resolution in the analysis process.

Narrow-band spectrograms were obtained from each FFR waveform using a 50 ms Hamming window to evaluate the spectral representation of the harmonics. In order to determine if there were between-group differences in the representation of the pitch-relevant harmonics in the FFR, the stimulus spectrum containing the first five harmonics was compared with that of the FFR response spectrum for each stimulus using a spectral correlation analysis confined to the region of maximal acceleration. The resulting correlation coefficient is expressed as a value between –1 and 1. Consistent with the pitch strength measurement, the stimulus time window varied in duration in order to include exactly five pitch periods.

Acknowledgments

Research supported by the National Institutes of Health R01 DC008549-01A1 (A.K.) and NIDCD predoctoral traineeship (G.B.). Thanks to Bruce Craig and Duncan Leaf for their assistance with statistical analysis (Purdue Department of Statistics).

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.bandl.2010.05.004.

References


