NeuroImage 120 (2015) 191–200

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Effects of language experience and stimulus context on the neural organization and categorical perception of speech

Gavin M. Bidelman [a,b,*], Chia-Cheng Lee [b]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152, USA
[b] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN 38152, USA

ARTICLE INFO

*Article history:*
Received 12 February 2015
Accepted 6 June 2015
Available online 3 July 2015

*Keywords:*
Auditory event-related potentials (ERP)
Experience-dependent plasticity
Categorical perception
Linguistic tones
Stimulus context effects
Mandarin Chinese

ABSTRACT

Categorical perception (CP) represents a fundamental process in converting continuous speech acoustics into invariant percepts. Using scalp-recorded event-related brain potentials (ERPs), we investigated how tone-language experience and stimulus context influence the CP for lexical tones—pitch patterns used by a majority of the world's languages to signal word meaning. Stimuli were vowel pairs overlaid with a high-level tone (T1) followed by a pitch continuum spanning between dipping (T3) and rising (T2) contours of the Mandarin tonal space. To vary context, T1 either preceded or followed the critical T2/T3 continuum. Behaviorally, native Chinese showed stronger CP as evident by their steeper, more dichotomous psychometric functions and faster identification of linguistic pitch patterns than native English-speaking controls. Stimulus context produced shifts in both groups' categorical boundary but was more exaggerated in native listeners. Analysis of source activity extracted from primary auditory cortex revealed overall stronger neural encoding of tone in Chinese compared to English, indicating experience-dependent plasticity in cortical pitch processing. More critically, "neurometric" functions derived from multidimensional scaling and clustering of source ERPs established: (i) early auditory cortical activity could accurately predict listeners' psychometric speech identification and contextual shifts in the perceptual boundary; (ii) neurometric profiles were organized more categorically in native speakers. Our data show that tone-language experience refines early auditory cortical brain representations so as to supply more faithful templates to neural mechanisms subserving lexical pitch categorization. We infer that contextual influence on the CP for tones is determined by language experience and the frequency of pitch patterns as they occur in listeners' native lexicon.



## Introduction

Categorical perception (CP) represents a fundamental aspect of human perception–cognition with a rich empirical history (for review, see Harnad, 1987). In speech, CP occurs when gradually morphed speech sounds along a large acoustic continuum are heard as belonging to one of only a few discrete phonetic classes or common identities (Liberman et al., 1967; Pisoni, 1973; Pisoni and Luce, 1987). Listeners treat sounds within a given category as perceptually similar despite their otherwise dissimilar acoustic characteristics. Tokens labeled with different phonetic identities are said to cross the categorical boundary, a psychological border where listeners' perception abruptly flips. Categorical speech boundaries emerge early in life (Eimas et al., 1971) and are further modified based on one's native tongue (Kuhl et al., 1992; Xu et al., 2006a). This suggests that the neural mechanisms underlying CP are malleable to the experiential effects of learning and language experience.

Neural correlates of CP have been identified in various neuroimaging studies (e.g., Bidelman, 2015; Binder et al., 2004; Chang et al., 2010; Zhang et al., 2011). In our recent ERP studies, we traced the functional neural chronometry of CP using event-related brain potentials (ERPs) evoked by a speech vowel continuum (Bidelman et al., 2013b). Comparisons between psychometric and "neurometric" identification functions—derived from their ERPs—revealed that listeners' perceptual CP for speech was well-predicted based on the magnitude of their underlying neural responses (time window of the P2 wave: ~150 ms). That is, brain activity tended to cluster according to phonetic rather the acoustic rules and closely mirrored their behavioral classification (Bidelman et al., 2013b) including predicting the steepness of their categorical boundary (Bidelman, 2015). These results suggest that the neural underpinnings of categorical speech perception emerge within the first few hundred milliseconds after sound enters the ear (see also Chang et al., 2010) and moreover, that the degree of CP listeners perceive is predicted by early cortical neural representations for speech. More recently, we have shown that auditory experience (e.g., musical training) can strengthen perceptual CP (e.g., faster and sharper classification) and that these behavioral enhancements are reflected in more dichotomous (i.e., categorical) neural activity (Bidelman and Alain, 2015).

* Corresponding author at: School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA. Fax: +1 901 525 1282.
*E-mail address:* g.bidelman@memphis.edu (G.M. Bidelman).

http://dx.doi.org/10.1016/j.neuroimage.2015.06.087
1053-8119/

Here, we investigated how other forms of auditory experience, namely tone-language expertise, confer similar experience-dependent neuroplasticity and influence categorical speech processing.

Pitch provides an optimal window to study language-dependent effects on CP as it is one of the most important information-bearing components of language. In particular, tone languages provide a unique opportunity for investigating the linguistic use of pitch in relation to categorical speech perception given their widespread use among the world's population. In these languages, pitch variations at the syllable or word level are lexically significant. Mandarin Chinese, for example, has four lexical tones: high-level tone *ma*$^1$ 'mother' [T1], rising tone *ma*$^2$ 'hemp' [T2], falling-rising tone *ma*$^3$ 'horse' [T3], falling tone *ma*$^4$ 'scold' [T4] (Yip, 2003). Cross-language comparisons between Mandarin and English speakers' perception of tonal continua demonstrate language-dependent enhancements in native listeners in the form of stronger, more dichotomous CP for the pitch patterns of the Mandarin tonal space (Peng et al., 2010; Xu et al., 2006a; Zheng et al., 2012). Additionally, Chinese listeners' more salient CP for even non-speech stimuli further suggests an experience-dependent enhancement to the categorization of pitch that is domain-general but also influenced by long-term categorical representations and short/long-term memory garnered through tone-language experience (Xu et al., 2006a). Consistent with behavioral studies, neuroimaging work demonstrates that linguistic pitch experience tunes both subcortical (Bidelman et al., 2011a, 2011b; Krishnan et al., 2010; Xu et al., 2006b) and cortical (Chandrasekaran et al., 2007b; Gandour et al., 1998, 2000) encoding of pitch. Presumably, tone-languages not only enhance neural pitch representations, but also act to warp or restrict the perceptual space near category boundaries to supply a more dichotomous decision when classifying lexical tones (e.g., Bidelman and Alain, 2015; Bidelman et al., 2014b; Iverson et al., 2003). While a number of previous studies have examined the neural correlates of categorical pitch perception in native tone-language listeners (e.g., mismatch negativity (MMN) studies: Xi et al., 2010; Zhang et al., 2012; Zheng et al., 2012), the absence of a control group (i.e., nonnative speakers) has obscured whether or not the observed effects were language-dependent and/or tone-specific, or more broadly, reflect a generalized CP for speech that is independent of language background (e.g., Bidelman et al., 2013b; Kuhl and Miller, 1975).

Additionally, while behavioral studies have demonstrated cross-language differences in the categorization of tone (e.g., Hallé et al., 2004; Lee et al., 1996; Xu et al., 2006a), surprisingly few studies have investigated how *contextual variation* in pitch affects its categorization (Francis and Ciocca, 2003). Among the handful of studies examining context effects, Xu (1997) indicated an asymmetry in tonal production whereby the effect of a preceding tone on the following tone, in terms of pitch height, was greater than in the reverse direction. Similarly, in disyllabic word perception, Ganong (1980) found that English listeners showed a perceptual bias and shift in their categorical boundary towards the real word "dash" for a "dash-tash" continuum. This suggests that stimulus context expands the mental category for expected or behaviorally relevant stimuli (McMurray et al., 2008). In terms of tone perception, Lin and Wang (1984) posited that when judging tones of a disyllable, listeners must compare the two pitch patterns in order to make their judgments; if the first syllable is short, listeners are unable to process the first tone before hearing the second and consequently report illusory percepts. For example, listeners sometimes misjudged T1–T2 as T3–T2 or T1–T2 as T2–T2 because they may have firstly defined the offset of the T2 contour as the highest pitch of the disyllable before completely processing the pitch of the first tone (Lin and Wang, 1984). More recently, Cao et al. (2012) found that Mandarin listeners categorized T2 and T3 differently depending on what tone preceded and followed a T2/T3 continuum. That is, listeners' perceptual boundary shifted in a stimulus-dependent manner. In contrast, nonnative listeners showed no signs of such contextual biases. Collectively, studies suggest that listeners' CP for pitch as well as their categorical boundaries vary as a function of surrounding stimulus context in a language-dependent manner. Under investigation here are the neural underpinnings of these context- and language-dependent influences on CP.

To this end, we recorded neuroelectric brain activity from native (Mandarin Chinese) and nonnative (English) listeners in response to a categorically perceived pitch continuum (T2–T3). Comparisons between language groups allowed us to elucidate potential cross-linguistic differences in the categorical neural organization and CP for tone. Based on our previous findings examining musical training and CP (Bidelman and Alain, 2015), we hypothesized that native Chinese listeners would show more dichotomous categorization for linguistic tones compared to nonnative English listeners given their extensive experience with linguistically-relevant pitch. More importantly, we predicted that their perceptual enhancements would be paralleled in stronger, more categorical neural representations for pitch as reflected in the ERPs. Secondly, we aimed to determine the influence of stimulus context on the categorical perception of linguistic tones. Thus, our stimuli included a third neutral tone (T1) that was presented immediately before or after the categorical continuum of interest. Extending previous behavioral studies (e.g., Cao et al., 2012), we expected native listeners to show more pronounced context effects than nonnative listeners as evident by more exaggerated shifts in both their psychometric and corresponding neurometric identification functions (derived from multidimensional scaling and clustering of ERPs).

## Materials & methods

### *Participants*

Twenty young adults participated in the experiment. Ten participants (5 females) served as the experimental group and were native speakers of Mandarin Chinese (hereafter referred to as C). An additional ten native English-speaking listeners (5 females) (hereafter referred to as E) served as the control group. Each participant completed a language history questionnaire (Li et al., 2006). Native speakers of Mandarin were considered sequential Mandarin–English bilinguals and were born and raised in mainland China. They had not received formal instruction in English before the age of 9 ($M \pm SD$; 10.5 ± 2.2 years). All grew up and earned their undergraduate degrees while in China and came to the USA for their graduate studies. Each currently spoke Mandarin at least 47.5% of their daily language use. Critically, none of the participants in the English group had prior experience learning Mandarin or a tone language of any kind. All participants had normal hearing, were right-handed [C: 93 ± 11.3; E: 89.5 ± 12.6%] (Oldfield, 1971), had obtained a similar level of education [C: 18.6 ± 3.1; E: 19.2 ± 1.7 years; $t(18) = 0.54$, $p = 0.59$], and were closely matched in age [C: 26.4 ± 4.6; E: 28.7 ± 2.8 years; $t(18) = 1.36$, $p = 0.19$]. None reported any history of neuro-psychiatric illness. Musical training is known to amplify the processing of linguistic tones (Bidelman et al., 2011a; Wong et al., 2007) and enhance categorical speech perception (Bidelman et al., 2014b). Hence, all participants were required to have minimal formal musical training (C: 0.2 ± 0.4 years; E: 1.2 ± 2.1 years; $t(18) = 0.54$, $p = 0.17$) and no instruction within the past five years. All were paid for their time and gave written informed consent in accordance with a protocol approved by the Institutional Review Board at the University of Memphis.

### *Stimuli*

#### *CP tonal continua*

Stimuli consisted of two tonal continua, designed to vary minimally acoustically in their pitch patterns, yet be perceived categorically (Fig. 1). Each set was constructed using five equally spaced pitch patterns, creating a tonal continuum between Mandarin Tone 2 (T2) and Tone 3 (T3). Specific pitch patterns were modeled after those reported by Cao et al. (2012). To examine contextual effects on the CP of linguistic

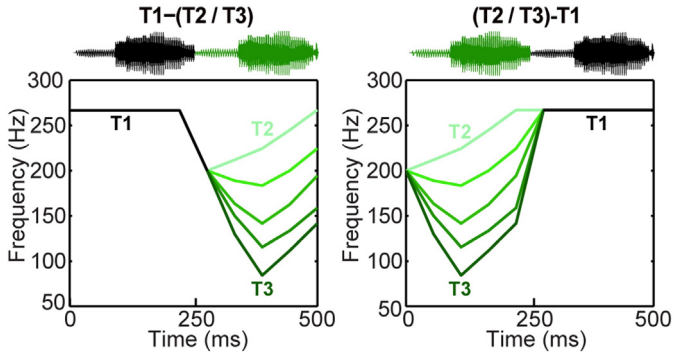


**Fig. 1.** Categorical pitch patterns used to probe contextual effects on lexical tone perception. Linguistic pitch patterns varied continuously between Mandarin Tone 2 (T2) and 3 (T3) but were perceived categorically (Cao et al., 2012). The T2/T3 continuum was presented in two different stimulus contexts where level Tone 1 (T1) preceded [T1-(T2/T3)] or followed the categorical pitch continuum [(T2/T3)-T1]. Each tonal pattern was overlaid onto a 500 ms disyllabic vowel sequence /mama/. Time waveforms (top) demarcate the amplitude profiles of the two vowel sequence.

tones, this T2/T3 continuum was presented in two different stimulus contexts. In the first context, T1 preceded the T2/T3 continuum [i.e., "T1-(T2/T3)" context]; in the second, it followed the categorical continuum [i.e., "(T2/T3)–T1"]. In both cases, the continua itself was otherwise identical; only its position relative to T1 differed between the two stimulus contexts. Each pair of tonal patterns was overlaid onto a 500 ms disyllabic vowel sequence /mama/ synthesized in PRAAT (Boersma and Weenink, 2013). It should be noted that the vowel sequence /mama/ with the current tonal patterns has no meaning in Mandarin. The length of each individual syllable (i.e., lexical tone segment) was 250 ms. In total, there were 10 stimulus conditions (2 contexts × 5 continuum steps). Stimulus presentation was controlled by MATLAB® 2014 (The MathWorks, Inc.) routed through a TDT RP2 interface (Tucker-Davis Technologies). Auditory stimuli were delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-2, Etymotic Research).

*Stimulus presentation and task*

During electrophysiological recordings, subjects performed a speeded categorization task of the tonal stimuli (e.g., Bidelman et al., 2013b, 2014a). Listeners heard 200 randomly ordered exemplars of each token and were asked to judge only the T2/T3 portion of the stimuli by labeling them with a binary response as quickly as possible ("T2" or "T3") via a keyboard button press (Cao et al., 2012). Ample practice was provided prior to starting the experimental run to ensure task familiarization. The two contexts were tested in separate blocks, counterbalanced across subjects. Following the participant's response, the interstimulus interval was jittered randomly between 400 and 600 ms (20-ms steps, rectangular distribution) and the next trial commenced. Listeners were offered a break after each block. Duration of testing including electrode preparation, testing, and breaks, lasted ~2 h.

*Electrophysiological recordings and preprocessing*

Participants sat comfortably in an electro-acoustically shielded booth to facilitate recording of neurophysiological responses. Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 locations around the scalp (Oostenveld and Praamstra, 2001) using standard procedures from our lab (e.g., Bidelman and Grall, 2014). Continuous EEGs were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online lowpass filter with a cutoff of 200 Hz. Responses were then stored to disk for offline analysis. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. During online acquisition, all electrodes were referenced to an additional sensor placed ~1 cm posterior to Cz. However, data were re-referenced off-line to a common average reference. Contact impedances were maintained ≤10 kΩ throughout the duration of the experiment.

Subsequent preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom routines coded in MATLAB. Ocular artifacts (saccades and blink artifacts) were corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom et al., 2004). The PCA decomposition provided a set of scalp projections (components) that best explained the average topography of the blink/saccadic artifacts measured by the EOG channels. The scalp projection of the first two PCA loadings was subtracted from the continuous EEG traces to nullify ocular contamination in the final ERPs. Cleaned EEGs were then digitally filtered (1–20 Hz; zero-phase filters), epoched (−200–1000 ms, where $t = 0$ was the onset of the stimulus), baseline-corrected to the pre-stimulus period, and subsequently averaged in the time domain to obtain ERPs for each stimulus condition per participant.

In our previous reports, robust categorical effects were observed in the auditory evoked potentials at frontocentral locations of the scalp (i.e., FCz electrode), indicative of bilateral sources in auditory cortex (Bidelman et al., 2013b, 2014a, 2014b). Nevertheless, neuronal sources of evoked potentials are a mixture of volume conducted potentials generated from a variety of active generators; "cross-talk" between adjacent sensor measurements precluded firm conclusions as to where in the brain these CP effects were actually generated. To more directly assess whether or not activity emitted from auditory cortex shows categorical processing, we performed a distributed source analysis (Bidelman and Dexter, 2015).

Source reconstruction was implemented in the MATLAB package Brainstorm (Tadel et al., 2011). We used a realistic, boundary element model (BEM) volume conductor (Fuchs et al., 1998, 2002) standardized to the MNI brain (Mazziotta et al., 1995). A BEM model was chosen as it is less prone to spatial errors than other head models (e.g., spherical conductor) (Fuchs et al., 2002). We used the well-established sLORETA inverse solution (Pascual-Marqui, 2002) to estimate the distributed neuronal current density underlying the recorded sensor data. This algorithm models the inverse solution as a large number of elementary dipole generators distributed over nodes on a mesh of the cortical surface. When constrained to neocortical layers, the aggregate strength of source activity can be projected spatiotemporally onto the neuroanatomy, akin to functional maps in fMRI. The resultant activation maps represent the transcranial current source density underlying the scalp-recorded potentials as seen from the cortical surface. We used the default settings in Brainstorm's implementation of sLORETA (Tadel et al., 2011). Time-courses of the sLORETA solution were then extracted within a predefined region of interest (ROI) situated in bilateral primary auditory cortex (A1) (i.e., Heschl's gyrus). ROI parcellation was based on anatomical segmentations (Fischl et al., 2004) and the OpenMEEG BEM head model (Gramfort et al., 2010) as implemented in Brainstorm (Tadel et al., 2011). Resulting source waveforms reflect the neural activity (current, measured in μAmm) as seen within the anatomical ROI (i.e., left/right A1). Visual inspection of the ERP scalp topographies revealed largely bilateral symmetric responses (i.e., no laterality) for both language groups (data not shown). Thus, left and right hemisphere A1 source waveforms were averaged to reduce the dimensionality of the data.

*Neurometric identification functions derived from cortical source ERPs*

We constructed "neurometric" identification functions derived from the source ERPs using a data-driven approach (i.e., multidimensional scaling and clustering of responses)—adopted and modified from techniques of Chang et al. (2010)—that implements the definition of CP (Bidelman and Alain, 2015; Bidelman et al., 2013b). The rationale behind this approach is that within category speech sounds are perceived as belonging to the same class and thus, should elicit similar neural activity patterns whereas across category tokens are heard as dissimilar and should elicit more divergent neural responses. In speech

perception, psychophysical differences are often explored using confusion matrices representing the perceptual dissimilarity between sound tokens. Analogous "neural dissimilarity" matrices (Bidelman and Alain, 2015; Bidelman et al., 2013b; Chang et al., 2010) were computed separately within a time window spanning 20-ms centered on the cortical P2 wave of each individuals' source ERPs. We restricted analysis to the P2 as derivations of neurometric functions from earlier sensory components (i.e., P1, N1) showed poor correspondence with behavior (data not shown). This is consistent with our previous studies which demonstrate that categorical neural organization in the ERPs of younger adults did not emerge prior to P2 (Bidelman et al., 2013b). Additionally, we have shown changes in CP with age (Bidelman et al., 2014a) and auditory experience, e.g., musical training (Bidelman and Alain, 2015; Bidelman et al., 2014b), are predicted based on neural activity in the P2 time window.

The P2 analysis window was adjusted according to the stimulus context so that it aligned with the timing of the critical T2/T3 continuum's onset and each subjects' P2 wave. For the T1-(T2/T3) context, P2 was taken as the positive deflection at an absolute latency of ~400 ms, that is, ~150 ms after the onset of the T2/T3 continuum in that condition (see Fig. 4, "P2b"). For the (T2/T3)-T1 context, the relevant P2 occurred much earlier (~150 ms) given that the onset of the continuum occurred at time = 0 (Fig. 4, "P2a"). Within the P2 analysis window for each context, the standardized Euclidean distance was computed between the source waveforms of all pairwise response combinations and used to construct a dissimilarity matrix (see Fig. 5, top panels). Each matrix cell quantifies the degree to which neuroelectric amplitudes within auditory cortex differs between a given pair of tonal stimuli.

Neural dissimilarity matrices were then submitted to a multidimensional scaling (MDS) analysis to visualize differences in neurophysiological responses across stimuli in a two-dimensional Euclidean space. MDS is a popular tool for examining perceptual dissimilarities between auditory stimuli (Borg and Groenen, 2005; Shepard, 1980); we used it here to similarly examine the dissimilarity of neural responses elicited by a speech continuum (Bidelman and Alain, 2015; Bidelman et al., 2013b). Graphically, points in MDS space represent empirical distances between brain responses akin to geographic distances between cities on a map. Within-category speech sounds are more difficult to discriminate and thus, elicit responses which are positioned closer together in MDS space; across-category tokens are easier to discriminate and hence, are well separated geometrically (Bidelman and Alain, 2015; Bidelman et al., 2013b; Chang et al., 2010). A stress value <0.1, which represents the reconstruction's "badness of fit", was obtained with a MDS solution of only two dimensions indicating an adequate fit to the data (Borg and Groenen, 2005).

We applied $k$-means clustering ($k = 2$) to the MDS solution to test whether ERPs grouped in a meaningful way that paralleled psychophysical classifications based solely on differences between the evoked activity generated by each lexical pitch pattern. The choice of two clusters was based on the a priori knowledge that perceptually, our stimuli fell into one of two distinct phonetic categories (i.e., "T2" and "T3") (cf. Bidelman and Alain, 2015; Bidelman et al., 2013b; Chang et al., 2010) and the fact that our behavioral task forced listeners to make a binary decision when labeling tokens. Objects falling within a given cluster were considered to be representatives of the same lexical tone identity (i.e., phonetic category) (see Fig. 5).

"Neurometric" identification functions were then constructed for each lexical tone category by computing the normalized distance between each of the individual responses (as represented in MDS space) and each of the two cluster means (representing a neural exemplar, or "template", for each tone category) (for details, see Bidelman et al., 2013b; Chang et al., 2010). Generated with respect to both cluster means, the resulting functions estimate how well neural activity evoked by each linguistic pitch pattern fits into one of the two discrete tonal categories (see Fig. 6). Comparisons between the degree to which Chinese- and English-speakers' neurometric functions paralleled their psychometric functions (i.e., via correlations) allowed us to assess potential language-dependent effects on CP and whether or not native listeners showed more pronounced categorical encoding/organization for pitch. Comparisons between contexts allowed us to verify that our data driven approach (i.e., "neurometric functions") is sensitive to, and able to predict, the contextual influences on the CP boundary observed behaviorally.

### *Behavioral data analysis*

Psychometric identification scores were fit with a two-parameter sigmoid function (Bidelman et al., 2011d, 2013b) as detailed in Xu et al. (2006a). We used standard logistic regression: $P = 1 / [1 + e^{-\beta 1(x - \beta 0)}]$, where $P$ is the proportion of trials identified as a given vowel, $x$, the step number along the stimulus continuum, and $\beta_0$ and $\beta_1$ the location and slope of the logistic fit estimated using an iterative, nonlinear least-squares regression procedure as implemented in MATLAB's 'nlinfit' function. Comparing parameters across groups and conditions revealed possible differences in the location and "steepness" (i.e., rate of change) of the categorical speech boundary as a function of language experience and stimulus context. Behavioral speech labeling speeds (i.e., reaction times; RTs) were computed as listeners' mean response latency across trials for a given condition. RTs outside 250–3500 ms were deemed outliers and excluded from further analysis (e.g., Bidelman and Alain, 2015; Bidelman et al., 2013b).

### *Statistical analysis*

Unless otherwise specified, two-way, mixed-model ANOVAs were conducted on the behavioral variables, i.e., CP location ($\beta_0$) and slope ($\beta_1$) parameters and RTs (SAS® 9.4, SAS Institute, Inc., Cary, NC, USA). Dependent measures were square-root transformed prior to analysis to improve homogeneity of variance assumptions necessary for parametric statistics. Group (2 levels; C, E) functioned as the between-subjects factor and tonal context [2 levels: T1-(T2/T3); (T2/T3)-T1] as the within-subjects factor; subjects nested within group served as a random factor. Tukey–Kramer multiple comparisons controlled Type I error inflation. Significance level was set at $\alpha = 0.05$.

## Results

### *Behavioral psychometric functions*

Behavioral speech identification functions are shown in Fig. 2 and CP slope and location measures in Fig. 3. Despite the continuous acoustic change in the stimuli, listeners heard a clear perceptual shift in the tonal category (T2 vs. T3) near the midpoint of the continuum which varied according to language experience and stimulus context (Fig. 3). In both groups, the overall location of the perceptual boundary shifted (leftward) according to the tonal context [sole main effect of context: $F_{1,\ 18} = 4.57$, $p = 0.046$]; listeners were more biased to hear "T3" in the (T2/T3)-T1 tonal context compared to the T1-(T2/T3) context. A priori comparisons by group revealed that the location of the CP boundary shifted with tonal context in Chinese listeners ($p = 0.028$) whereas it was invariant in the English group ($p = 0.57$) (cf. Figs. 3B and E). Comparisons of the slope of listeners' psychometric functions (Figs. 3A and D) revealed a main effect of group [$F_{1,\ 18} = 9.12$, $p = 0.0074$]. That is, Chinese listeners obtained steeper classification functions than their English counterparts across the board.

By context, an ANOVA (tonal stimulus, 5 levels, group: 2 levels) on speech labeling speeds (RTs) revealed a group × stimulus interaction in the T1-(T2/T3) context [$F_{4,\ 72} = 2.55$, $p = 0.046$] (Fig. 3C). The significant interaction indicates a differential pattern of labeling speeds between language groups. Post-hoc contrasts revealed that Chinese listeners were slower at classifying speech tokens near the CP boundary (token 3) relative to others in the continuum [$t_{72} = 5.71$, $p < 0.0001$]. In

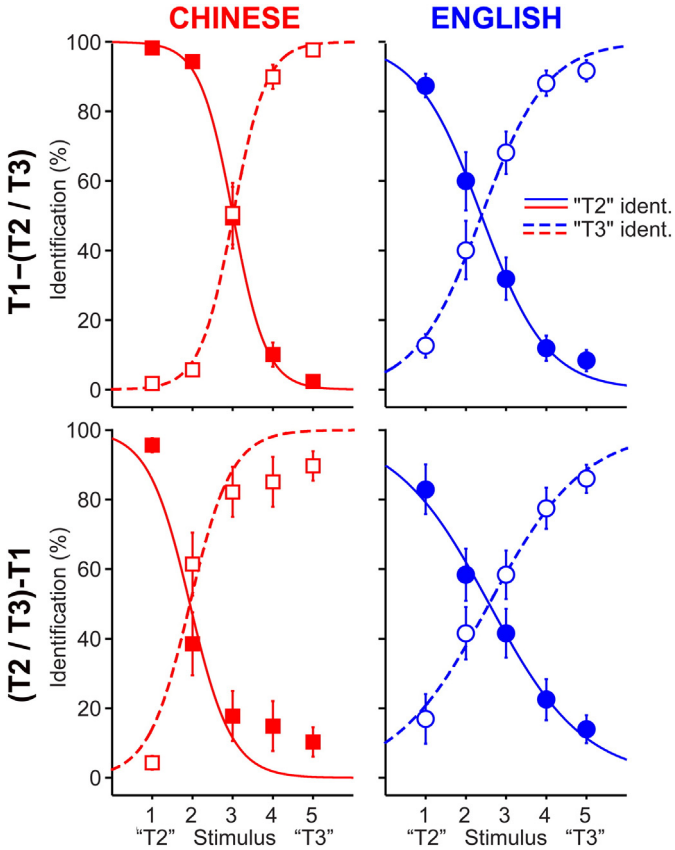


**Fig. 2.** Psychometric tone identification for Chinese and English listeners as a function of stimulus context. Despite the continuous acoustic change in the stimuli, listeners heard a clear perceptual shift in the tonal category (T2 vs. T3) near the midpoint of the continuum that varied according to language experience and stimulus context. The overall location of the perceptual boundary shifted leftward in the (T2/T3)-T1 compared to the T1-(T2/T3) context, particularly in the Chinese group, demonstrating a contextual effect on the categorization of lexical tones. Chinese generally showed sharper identification functions than English listeners indicating more dichotomous internalized representations for pitch patterns. Error bars = s.e.m.

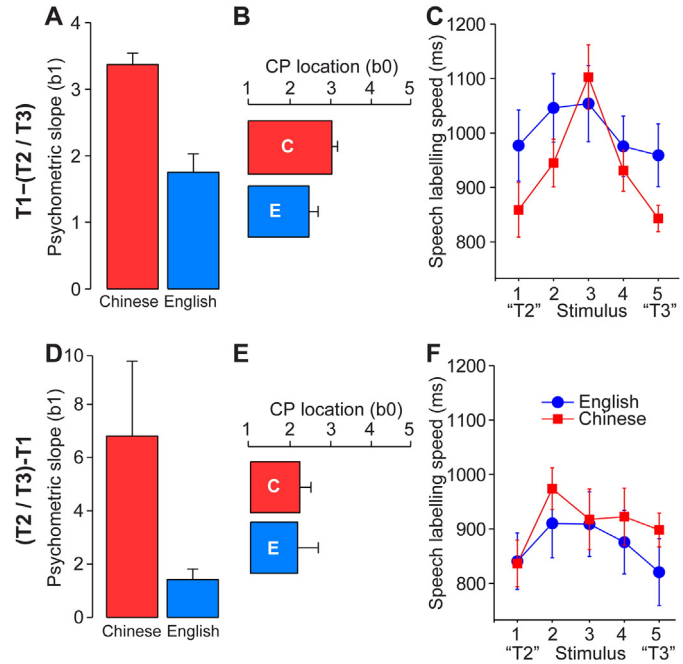


**Fig. 3.** Cross-language differences in categorical perception. Comparison of the "steepness" (A) and location (B) of psychometric identification functions per language group for the T1-(T2/T3) tonal context. (C) Speech labeling speeds for tone identification. Chinese show steeper identification functions than English speakers, indicating greater perceptual distinction between pitch categories. Chinese listeners are slower to label sounds near the categorical boundary (stimulus 3)—indicative of CP—whereas English classification speeds are invariant across the continuum. (D–F) same as panel A–C but for the (T2/T3)-T1 context. Chinese obtain steeper identification than their English-speaking peers but with a leftward shift in the perceptual boundary and less categorical labeling speeds. Error bars = s.e.m.

contrast, English listeners RTs were invariant across the continuum [$t_{72} = 1.74$, $p = 0.08$]. The relative slowing of speech labeling speeds near the CP boundary is consistent with previous reports examining speeded vowel classification (Bidelman et al., 2013b, 2014b; Pisoni and Tash, 1974). The fact that this pattern is only observed in the Chinese further indicates that the T1-(T2/T3) context is perceived more categorically in native compared to nonnative listeners (cf. Bidelman et al., 2014b).

For the (T2/T3)-T1 context, the ANOVA revealed a sole main effect of stimulus on behavioral RTs [$F_{4, 72} = 2.86$, $p = 0.0294$] (Fig. 3F). Follow-up contrasts failed to reveal reliable differences across stimuli in either group. Thus, unlike the T1-(T2/T3) context, there were no language-dependent differences in speech labeling speeds for the (T2/T3)-T1 context. Collectively, behavioral results indicate that (i) Chinese listeners have stronger categorical perception for linguistic pitch patterns than English listeners and (ii) CP of lexical tones varies with stimulus context, being much stronger in native speakers when the categorical continuum follows compared to when it precedes a precursor tone, i.e., T1-(T2/T3) vs. (T2/T3)-T1 context, respectively.

*Cortical ERPs to CP tones*

Grand average cortical source responses for each language group and stimulus context are shown in Fig. 4. Obligatory waves of the auditory cortical ERPs appeared in the first ~150–200 ms following the time-locking stimulus. Early components (labeled N1a, P2a) reflect the encoding of the initial syllable. Following the onset of the second lexical tone, another N1–P2 signature is observed (labeled N1b, P2b). P2 amplitudes were weaker in the T1-(T2/T3) context due to neural habituation and response overlap with the first tone; P2b was similar in both amplitude [$F_{1, 72} = 0.84$, $p = 0.37$] and latency [$F_{1, 72} = 1.38$, $p = 0.25$] between groups. However, Chinese showed stronger [$F_{1, 72} = 4.38$, $p = 0.05$] and earlier (~10 ms) [$F_{1, 72} = 5.99$, $p = 0.02$] P2a responses to lexical tones than the English group for the (T2/T3)-T1 context, consistent with the well-known experience-dependent enhancements of pitch processing in native tone-language speakers (e.g., Bidelman et al., 2011a, 2011b; Chandrasekaran et al., 2007b, 2009a). These findings corroborate the notion that tone-language speakers generally have enhanced neural encoding of lexical pitch compared to non-native listeners and is largely expected based on prior work (e.g., Bidelman et al., 2011a, 2011b, 2013a; Chandrasekaran et al., 2007b; Giuliano et al., 2011; Krishnan et al., 2010). In addition to these simple group differences, we aimed to identify if Mandarin speakers showed more categorical neural organization in their neural response to tone than English-speaking individuals.

*Neurometric classification derived from cortical ERPs*

Using differences between P2a/P2b current source responses, we constructed neural dissimilarity matrices, analogous to perceptual confusion matrices, to quantify the degree to which each group's brain activity could differentiate linguistic pitches and evaluate how neural organization for tone changed with stimulus context (e.g., see Fig. 5a in Bidelman and Alain, 2015). MDS applied to dissimilarity scores provides a visualization of response dissimilarities in a common Euclidean space; distances between objects quantify the magnitude of neural response dissimilarity (Fig. 5). MDS "maps" showed that ERPs to within-category tonal stimuli elicited similar patterns of neural activity and

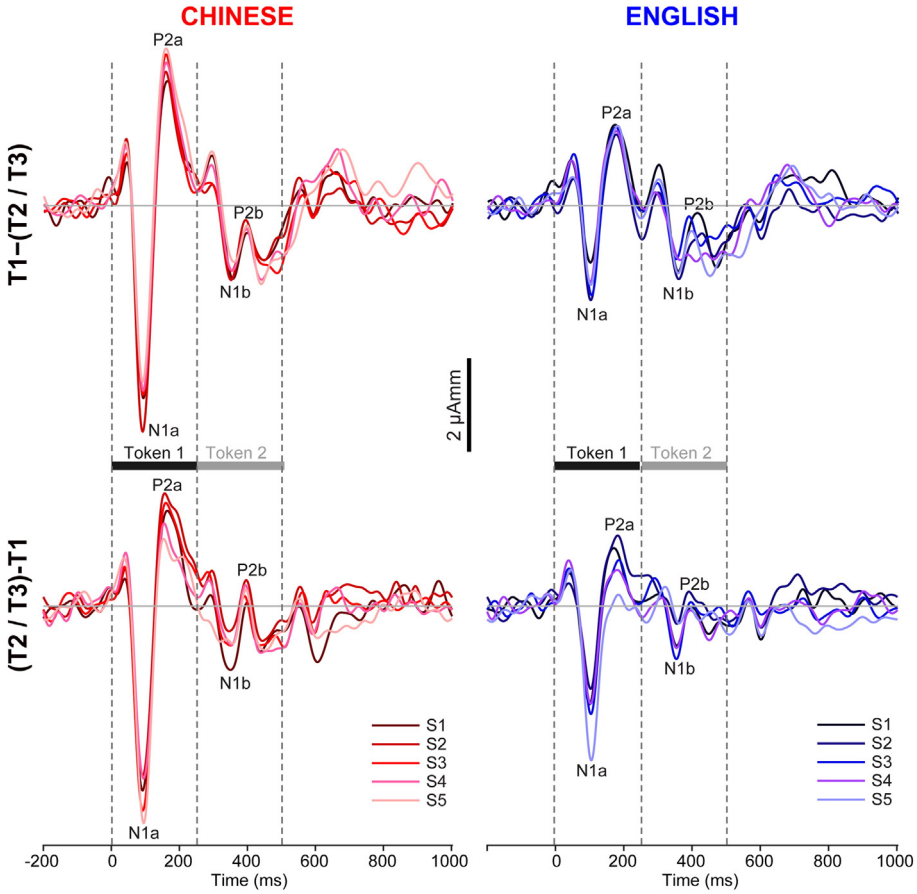


**Fig. 4.** Cortical source waveforms to CP tonal stimuli as a function of language experience and stimulus context. Traces represent the grand average neuroelectric activity (measured as current source amplitudes; μAmm) extracted from the primary auditory cortices via distributed source analysis. "Token 1" and "Token 2" (dotted lines) demarcate the two syllable segments of the disyllabic pair (see Fig. 1). Obligatory waves of the auditory cortical response appear in the first ~200 ms following the time-locking stimulus (N1a, P2a) and again following the onset of the second lexical tone (N1b, P2b). Chinese show stronger encoding of lexical tones than English speakers. Note the modulations in the P2 waves (P2a and P2b) across the stimulus continua which mark the time window for deriving neurometric identification functions (i.e., Fig. 5). S1-S5, continuum stimulus token.

appeared with closer proximity in geometric space; across-category tokens elicited divergent activity and were mapped farther apart. Clustering performed on MDS solutions revealed that cortical responses generated across the tonal continua could be meaningfully segregated into two distinct groupings (i.e., "T2" and "T3" clusters) mimicking the two phonetic classes of lexical tone heard by listeners (Fig. 5). Chinese MDS maps revealed that speech sounds clustered in a more consistent manner which paralleled perception. For example, tokens 1–3 grouped near one another but were remote from tokens 4–5. In contrast, English listeners' maps revealed patterns of miss-classifications where perceptually similar tones were erroneously assigned to opposing clusters (e.g., token 1 and token 4, Fig. 5A, right panels).

Neurometric identification functions were derived for each language group and stimulus context using the distance between evoked responses (as represented in MDS space) elicited by each tonal stimuli (i.e., Fig. 5) and the two cluster means (representing a neural "template" for each tone category), i.e., "T2" vs. "T3". Neurometric functions provide estimates of how well each group's pattern of neural activity evoked by each pitch pattern fit into one of the two discrete tonal categories. At the group level, Chinese neural classification functions for the T1-(T2/T3) context were strikingly similar to behavioral identification scores and closely mirrored their psychometric counterparts (Fig. 6A) ["T2" brain-behavior correlation: $r = 0.97$, $p = 0.0061$; "T3": $r = 0.98$, $p = 0.0047$]. Contrastively, English-speakers' neural responses were less reliable in predicting their behavioral CP ["T2": $r = 0.91$, $p = 0.01$; "T3": $r = 0.86$, $p = 0.07$]. Similarly, for the (T2/T3)-T1 context, neural identification functions showed a slight leftward shift in the location of the CP boundary, mirroring the leftward bias in listeners' perception (e.g., Fig. 2). Correlations between neuro- and psycho-metric functions were less robust for this context for both Chinese ["T2": $r = 0.74$, $p = 0.15$; "T3": $r = 0.87$, $p = 0.05$] and English ["T2": $r = 0.96$, $p = 0.0068$; "T3": $r = 0.04$, $p = 0.95$] listeners.

These group-level correlations were confirmed at the individual-level. Critically, brain-behavior correlations computed separately for each participant revealed stronger correspondence between neurometric and psychometric functions in Chinese listeners (Fig. 6C). An ANOVA revealed a significant main effect of group on brain-behavior correlations [$F_{1,\ 18} = 5.56$, $p = 0.029$]. The sole main effect of group (with no context effect or group × context interaction) indicates that Chinese listeners' neural responses showed closer correspondence with behavior across the board. Collectively, our findings demonstrate that cortical neural activity contains sufficient information (i) to predict listeners' categorical perception of linguistic tones (cf. speech vowels: Bidelman and Alain, 2015; Bidelman et al., 2013b) and (ii) cross-language enhancements and contextual influences on the categorical processing of linguistic pitch.

## Discussion

We investigated cross-language and contextual stimulus effects on the cortical encoding of linguistic pitch patterns. Findings of the current study relate to three primary observations: (1) the sequential order of linguistic tones modulates listeners' categorical perception (i.e., identification) of lexical pitch patterns and these contextual effects are more prominent in native Chinese listeners; (2) there is higher coupling between brain and behavioral responses to linguistic pitch in tone- vs. non-tone

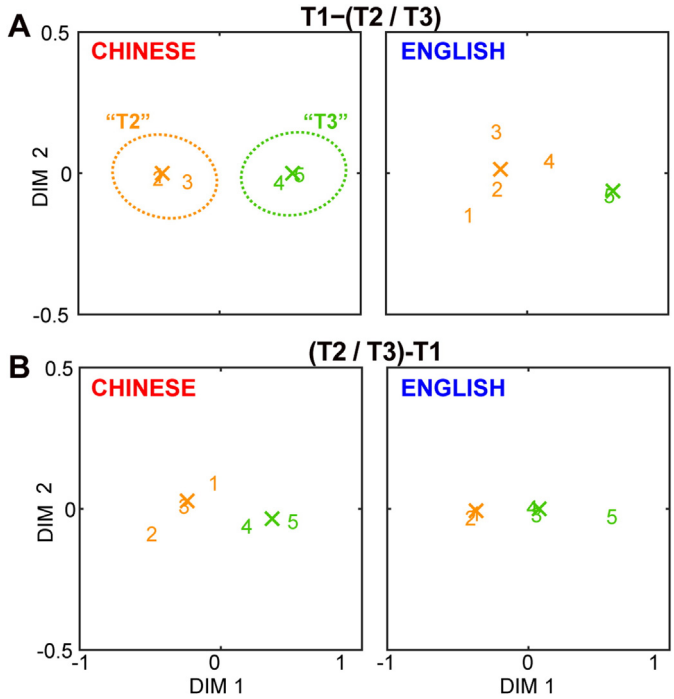


**Fig. 5.** Neural organization for pitch is more categorical in native compared to nonnative tone-language listeners. Multidimensional scaling (MDS) solution of neural dissimilarity matrices quantify the Euclidean distance between source activity across the tonal continua for each group and stimulus context. (A) T1-(T2/T3) context; (B) (T2/T3)-T1 context. Stimuli eliciting similar neural responses appear closer in MDS space, akin to adjacent cities on a map. Two meaningful clusters emerge (orange = "T2" cluster; green = "T3" cluster; 'x' = cluster centroids) mimicking the two lexical categories of the stimulus set. Examples of the T2 and T3 classification regions based on k-means clustering of the MDS space are shown as dotted ellipsoids in the upper left panel.

language listeners; and (3) there is a stronger, more categorical neural organization for pitch in primary auditory cortex for native Chinese compared to their English speaking counterparts. Collectively, our findings indicate that both stimulus context and language experience modulate the early auditory cortical encoding and behavioral categorization of pitch.

*Tone language experience strengthens categorical neural organization for pitch*

In the present study, we found overall enhanced cortical responses to linguistic pitch patterns in Chinese compared to English-speaking listeners. More robust cortical encoding of pitch in tone-language listeners is consistent with recent neurophysiological studies which have demonstrated that both cortical (Chandrasekaran et al., 2007a; Gandour et al., 2000; Giuliano et al., 2011) and subcortical (Bidelman et al., 2011a, 2011b, 2011c; Krishnan et al., 2010, 2011) auditory processing is enhanced in Chinese listeners in response to complex pitched stimuli. These findings further bolster the notion that long-term language experience produces experience-dependent changes in auditory brain function and improves the psychophysiological processing of tone. In the case of tonal languages, extensive experience with linguistic pitch shapes auditory processing in accordance with the functional demands of the Mandarin tonal space, conferring a listener with higher sensitivity and fidelity to pitch-relevant information (e.g., Bidelman et al., 2011a). Our findings also corroborate previous MMN studies which have shown larger mismatch responses for native phonetic deviants, but not nonnative or within-category speech contrasts (e.g., Dehaene-Lambertz, 1997; Dorman, 1974; Phillips et al., 2000) but see (Maiste et al., 1995). Similar categorical encoding has been reported in speakers of tonal languages, who show stronger MMNs for between- relative to within-category linguistic tones (Xi et al., 2010; Zhang et al., 2012; Zheng et al., 2012). Prior studies have thus shown that language experience can at least sensitize the brain's discrimination of native and nonnative speech categories.

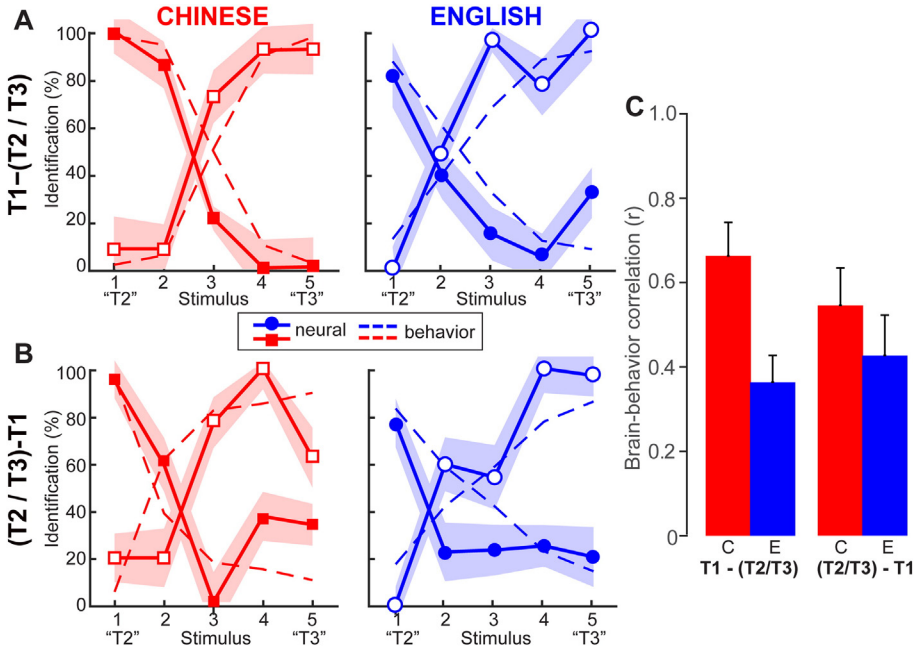


**Fig. 6.** Cross-language and contextual comparison of neurometric (solid) and psychometric (dotted) categorical identification functions. (A–B) Neurometric functions derived by computing the Euclidean distance between neural responses, as represented in MDS space (see Fig. 5) and the two cluster means, representing the prototypical "T2" and "T3" categories. Neural functions closely mirror behavior in Chinese listeners but are much less faithful in the English group. Across stimulus contexts (A vs. B), the location of the neurometric CP boundary shifts slightly leftward, paralleling the perceptual bias with context observed in listeners' psychometric functions (see Fig. 2). (C) Brain-behavior correlations for each group and stimulus context. Chinese showed closer correspondence between their psychometric and neurometric classification functions than English, indicating stronger categorical neural organization for linguistic pitch patterns. Error bars and shading = s.e.m.

Our findings further extend these previous studies by demonstrating that tone-language speakers actually have stronger, more dichotomous neural representations of linguistically relevant pitch contours which mirrors their more categorical perception of tone (present study; Peng et al., 2010; Xu et al., 2006a). Comparison between "neurometric" and psychometric tone identification revealed a stronger categorical organization of tone in Chinese listeners, i.e., neural responses more closely mirrored perception (Figs. 5–6). Indeed, across both stimulus contexts, Chinese showed a closer correspondence between brain and behavioral responses than English listeners. These findings imply a differential pattern of speech processing between language groups whereby native listeners' neural code carries more behaviorally-relevant information of the speech signal than nonnative listeners (at least with regard to pitch). They also complement our recent studies demonstrating that other forms of intensive pitch experience (e.g., musical training) similarly yield improved CP at both behavioral and neural levels (Bidelman and Alain, 2015; Bidelman et al., 2014b). We infer that tone-language expertise, as with musicianship (Bidelman and Alain, 2015), acts to warp or restrict the perceptual space near categorical pitch boundaries, supplying a more dichotomous decision when classifying sound objects. A more categorical neural organization for lexical tones—a direct consequence of speaking a tone-language—may account for the sharper, more dichotomous, and faster classification of linguistically relevant pitch contours we find in Chinese participants behaviorally (Fig. 3). There is some indication that native speaker's categorization of pitch is domain general, that is, is observed for both speech and non-speech stimuli that contain linguistically relevant pitch (Chandrasekaran et al., 2009b; Xu et al., 2006a). Of interest to future studies would be to examine if Chinese listeners' CP benefits observed here generalize to non-speech stimuli (e.g., hummed pitch patterns).

Interestingly, categorical encoding was observed in the timeframe of the P2 complex, early auditory cortical components thought to reflect synchronized neural activity from thalamic and early auditory cortical generators (Näätänen and Picton, 1987; Picton et al., 1999). Categorical speech coding likely depends on both the time course (Bidelman et al., 2013b; Toscano et al., 2010) and a variety of spatiotemporally distributed brain areas, including frontal and middle temporal gyri (Myers et al., 2009; Zhang et al., 2011). Yet, our results examining cortical source activity demonstrates that primary auditory cortex is capable of coding higher-order categorical information within the timeframe of the P2 (~150–175 ms after the phonetic continuum). Recent intracranial recordings have implicated aspects of posterior superior temporal gyrus in categorical (i.e., phonetic) speech processing (Chang et al., 2010). Using comparable analysis techniques, our far-field data extend these findings by demonstrating CP organization emerges as early as primary auditory regions (A1). They also challenge the notion that early auditory cortical activity is simply an exogenous reflection of the stimulus. Rather, the fact that CP for linguistic pitch emerges by the P2 wave (150–200 ms) provides convincing evidence that the categorical (i.e., phonetic) representation of speech emerges within the first few hundred milliseconds after the onset of sound (Bidelman and Alain, 2015; Bidelman et al., 2013b; Chang et al., 2010).

Categorical coding in this early time window also contrasts findings from Zheng et al. (2014), who suggested that the classification of lexical tones is apparent only in more endogenous, attention-dependent neural responses (e.g., P300). However, the majority of previous ERP studies on the CP of lexical tones have employed the MMN responses recorded in passive oddball paradigms (Xi et al., 2010; Zhang et al., 2012; Zheng et al., 2012). Taken together, it seems that categorical neural organization emerges only when listeners actively attend and classify sounds (Bidelman et al., 2013b) (but see Chang et al., 2010, for an example of passive CP). As such, passively evoked MMNs may not capture neural correlates of CP (e.g., Xi et al., 2010; Zhang et al., 2012; Zheng et al., 2012). Regardless of whether or not attention is a necessary requirement for the categorical neural coding we observe, we note CP effects are fast, and influence neural representations for pitch within 150–200 ms of sound entering the ear.

*Context effects in the categorical perception of lexical tone*

A basic tenant of perception is that humans form equivalence classes by assigning similar objects to the same categorical membership (Goldstone and Hendrickson, 2010). An extreme view of CP is that once established, internalized speech prototypes (i.e., equivalence classes) are invariant to superficial stimulus manipulation or context (Liberman et al., 1957). This strong position predicts that stimulus context should have little effect on CP. That is, sound elements that precede or follow a stimulus continuum should not influence its categorization or location of the perceptual boundary. On the contrary, context effects have been observed in CP for speech (Holt and Lotto, 2010). Indeed, the location of the CP boundary can shift dependent on surrounding acoustic characteristics (e.g., Ganong, 1980; Holt and Lotto, 2010; Pisoni, 1975, current study), suggesting that phonetic categories are flexible and depend partially on other adjacent signals (Repp and Liberman, 1987).

While context effects in CP have been established for consonant and vowel sounds, we are aware of only a handful of studies that have examined how contextual variation modulates the CP for lexical tones (Cao et al., 2012; Cao and Wang, 2011; Francis and Ciocca, 2003; Wong and Diehl, 2003) and no study that has examined cross-language differences in these phenomena. Here, we show that the order of linguistic tones strongly influences their CP. We found that Chinese listeners showed stronger, more categorical perception when the tonal continuum was preceded by a subsequent tone, i.e., T1-(T2/T3) context (Fig. 3). This was evident in the symmetry of identification at the perceptual boundary and speech labeling speeds, which showed the stereotyped slowing of responses near the categorical decision boundary (cf. Bidelman et al., 2013b, 2014b; Pisoni and Tash, 1974). In contrast, listeners showed less categorical perception when classifying an identical tonal continuum when it was followed by a subsequent tone [i.e., (T2/T3)-T1 context]. Additionally, they showed a perceptual bias for hearing T3 despite equidistant stimulus spacing (see Fig. 2). More critically, perceptual responses were closely paralleled in listeners' "neurometric" identification functions which showed a similar shift in the CP boundary. Cross-language differences in the (T2/T3)-T1 context suggest that Chinese listeners may have perceived each disyllable as a unit (i.e., words but unnatural combinations), whereas English listeners perceived each disyllable as two unrelated syllables. Regardless, these findings provide clear evidence that the context in which lexical tones occur influences their neural organization and subsequent behavioral categorization. Moreover, they show that stimulus context can expand the category of the expected stimulus (McMurray et al., 2008), as reflected in both neural and behavioral responses.

But how do we account for the psychophysiological bias in the (T2/T3)-T1 context? Contextual tonal variation could be related to the frequency of linguistic pitch combinations used in daily communication. If certain tonal combinations are utilized more often, listeners might be more sensitized to particular orderings of disyllables compared to others. To explicitly test this hypothesis, we collected frequency data for 3000 bigrams (sequences of two adjacent characters) in news ($n = 1500$) and fiction ($n = 1500$) excerpts from a searchable, online Chinese corpus of text (Da, 2004). We used the default parameters (bigram frequency ≥ 50 and mutual information value ≥ 3.5) to retrieve bigrams and then a native speaker (second author) judged whether the 3000 pairs were meaningful words. Consequently, we identified 1317 meaningful disyllables from the news database and 1338 disyllables from the fiction database. We found that the sequence T1-(T2/T3) occurred more frequently (news: $n = 119$; fiction: $n = 123$) than (T2/T3)-T1 combination (news: $n = 90$; fiction: $n = 94$). These ratings were confirmed by a second native speaker of Mandarin naïve to the purposes of the experiment (inter-reliability correlation:

$r = 0.97$). Speculatively, we suggest that our listeners may have been more aware of and required less time to process the T1-(T2/T3) combination given its preponderance in the Chinese lexicon. As implied by our preliminary lexical analysis, it is possible that such frequency effects explain why native listeners in the current study heard the T1-(T2/T3) context more categorically and were faster at classifying tokens than in the (T2/T3)-T1 context. Nevertheless, the limited size of the corpus examined here may not provide sufficient coverage of Mandarin's bigram statistics. Future lexical analyses employing more extensive written and spoken word corpuses are needed to fully test the role of behavioral relevance and stimulus experience on CP.

## Conclusion

Collectively, our data lead us to infer that long-term tone-language experience tunes early auditory cortical mechanisms responsible for the categorization of pitch. Moreover, these experience-dependent effects are malleable to stimulus context as pitch identification varies according to the order and context of the surrounding acoustic signal. Language- and stimulus-dependent factors in CP interact and seem to depend on the frequency with which tones naturally occur in a listener's native lexicon; more frequent pitch contexts are perceived more categorically than those which occur less often.

## Acknowledgments

Portions of this work were supported by grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology Foundation (AAAF) awarded to G.M.B. Correspondence and requests for materials should be addressed to G.M.B.

**Conflict of interest**

The authors declare no competing financial interests.

## References

Bidelman, G.M., 2015. Induced neural beta oscillations predict categorical speech perception abilities. Brain Lang. 141, 62–69.
Bidelman, G.M., Alain, C., 2015. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. J. Neurosci. 35, 1240–1249.
Bidelman, G.M., Dexter, L., 2015. Bilinguals at the "cocktail party": dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. Brain Lang. 143, 32–41.
Bidelman, G.M., Grall, J., 2014. Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex. NeuroImage 101, 204–214.
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011a. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. 23, 425–434.
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011b. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77, 1–10.
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011c. Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. Neurosci. Lett. 503, 203–207.
Bidelman, G.M., Krishnan, A., Gandour, J.T., 2011d. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. Eur. J. Neurosci. 33, 530–538.
Bidelman, G.M., Hutka, S., Moreno, S., 2013a. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PLoS One 8, e60676.
Bidelman, G.M., Moreno, S., Alain, C., 2013b. Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage 79, 201–212.
Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014a. Age-related changes in the subcortical–cortical encoding and categorical perception of speech. Neurobiol. Aging 35, 2526–2540.
Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014b. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur. J. Neurosci. 40, 2662–2673.
Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.
Boersma, P., Weenink, D., 2013. Praat: doing phonetics by computer (Version 5.3.51) [computer program] Retrieved September 1, 2013, from http://www.praat.org.
Borg, I., Groenen, P.J.F., 2005. Modern Multidimensional Scaling: Theory and Applications. 2nd ed. Springer, New York.
Cao, W., Wang, Q., 2011. Comparative Study on Chinese Tone Perception: a Report on Falling f0 Contours. ICPhS XVII, Hong Kong. pp. 404–407.
Cao, W., Ota, Y., Jiang, H., Liu, L., Zhang, J., 2012. Identification of Mandarin tone-2 and tone-3 in disyllabic contexts by Chinese natives and Japanese students. TAL 2012 Third International Symposium on Tonal Aspects of Languages, Nanjing, China, pp. 1–5.
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2007a. Experience-dependent neural plasticity is sensitive to shape of pitch contours. Neuroreport 18, 1963–1967.
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2007b. Mismatch negativity to pitch contours is influenced by language experience. Brain Res. 1128, 148–156.
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009a. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. Brain Lang. 108, 1–9.
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009b. Sensory processing of linguistic pitch as reflected by the mismatch negativity. Ear Hear. 30, 552–558.
Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432.
Da, J., 2004. Chinese text computing. http://lingua.mtsu.edu/chinese-computing.
Dehaene-Lambertz, G., 1997. Electrophysiological correlates of categorical phoneme perception in adults. Neuroreport 8, 919–924.
Dorman, M.F., 1974. Auditory evoked potential correlates of speech sounds discrimination. Percept. Psychophys. 15, 215–220.
Eimas, P.D., Siqueland, E.R., Jusczyk, P., Vigorito, J., 1971. Speech perception in infants. Science 171, 303–306.
Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.
Francis, A.L., Ciocca, V., 2003. Stimulus presentation order and the perception of lexical tones in Cantonese. J. Acoust. Soc. Am. 114, 1611–1621.
Fuchs, M., Drenckhahn, R., Wischmann, H.-A., Wagner, M., 1998. An improved boundary element method for realistic volume-conductor modeling. IEEE Trans. Biomed. Eng. 45, 980–997.
Fuchs, M., Kastner, J., Wagner, M., Hawes, S., Ebersole, J., 2002. A standardized boundary element method volume conductor model. Clin. Neurophysiol. 113, 702–712.
Gandour, J.T., Wong, D., Hutchins, G., 1998. Pitch processing in the human brain is influenced by language experience. Neuroreport 9, 2115–2119.
Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., Hutchins, G.D., 2000. A crosslinguistic PET study of tone perception. J. Cogn. Neurosci. 12, 207–222.
Ganong, W.F., 1980. Phonetic categorization in auditory word perception. J. Exp. Psychol. Hum. Percept. Perform. 6, 110–125.
Giuliano, R.J., Pfordresher, P.Q., Stanley, E.M., Narayana, S., Wicha, N.Y., 2011. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Front. Psychol. 2, 146.
Goldstone, R.L., Hendrickson, A.T., 2010. Categorical perception. Wiley Interdiscip. Rev. Cogn. Sci. 1, 69–78.
Gramfort, A., Papadopoulo, T., Olivi, E., Clerc, M., 2010. OpenMEEG: opensource software for quasistatic bioelectromagnetics. Biomed. Eng. Online 45.
Hallé, P.A., Chang, Y.-C., Best, C.T., 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. J. Phon. 32, 291–453.
Harnad, S.R., 1987. Categorical Perception: the Groundwork of Cognition. Cambridge University Press, New York.
Holt, L.L., Lotto, A.J., 2010. Speech perception as categorization. Atten. Percept. Psychophys. 72, 1218–1227.
Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, 47–57.
Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguistics 23, 81–95.
Krishnan, A., Gandour, J.T., Ananthakrishnan, S., Bidelman, G.M., Smalt, C.J., 2011. Functional ear (a)symmetry in brainstem neural activity relevant to encoding of voice pitch: a precursor for hemispheric specialization? Brain Lang. 119, 226–231.
Kuhl, P.K., Miller, J.D., 1975. Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. Science 190, 69–72.
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.
Lee, Y.-S., Vakoch, D., Wurm, L., 1996. Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. J. Psycholinguist. Res. 25, 527–542.
Li, P., Sepanski, S., Zhao, X., 2006. Language history questionnaire: a web-based interface for bilingual research. Behav. Res. Methods 38, 202–210.
Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phonemic boundaries. J. Exp. Psychol. 54, 358–368.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74, 431–461.
Lin, T., Wang, W.S.-Y., 1984. Perception of tones. Bull. Chin. Linguist. 2, 59–69.
Maiste, A.C., Wiens, A.S., Hunt, M.J., Scherg, M., Picton, T.W., 1995. Event-related potentials and the categorical perception of speech sounds. Ear Hear. 16, 68–90.
Mazziotta, J.C., Toga, A.W., Evans, A., Lancaster, J.L., Fox, P.T., 1995. A probabilistic atlas of the human brain: theory and rational for its development. NeuroImage 2, 89–101.
McMurray, B., Dennhardt, J.L., Struck-Marcell, A., 2008. Context effects on musical chord categorization: different forms of top-down feedback in speech and music? Cogn. Sci. 32, 893–920.
Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol. Sci. 20, 895–903.
Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112, 713–719.
Pascual-Marqui, R.D., 2002. Standardized low resolution brain electromagnetic tomography (sLORETA). Methods Find. Exp. Clin. Pharmacol. 24D, 5–12.
Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., Wang, W.S.Y., 2010. The influence of language experience on categorical perception of pitch contours. J. Phon. 38, 616–624.
Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., Roberts, T., 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cogn. Neurosci. 12, 1038–1055.
Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., Trujillo, N.J., 1999. Intracerebral sources of human auditory-evoked potentials. Audiol. Neuro Otol. 4, 64–79.
Pisoni, D.B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13, 253–260.
Pisoni, D.B., 1975. Auditory short-term memory and vowel perception. Mem. Cogn. 3, 7–18.
Pisoni, D.B., Luce, P.A., 1987. Acoustic-phonetic representations in word recognition. Cognition 25, 21–52.
Pisoni, D.B., Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. Percept. Psychophys. 15, 285–290.
Repp, B.H., Liberman, A.M., 1987. Phonetic category boundaries are flexible. In: Harnad, S.R. (Ed.), Categorical Perception: the Groundwork of Cognition. Cambridge University Press, New York, pp. 89–112.
Shepard, R.N., 1980. Multidimensional scaling, tree-fitting and clustering. Science 210, 390–398.
Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 1–13.
Toscano, J.C., McMurray, B., Dennhardt, J., Luck, S.J., 2010. Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. Psychol. Sci. 21, 1532–1540.
Wallstrom, G.L., Kass, R.E., Miller, A., Cohn, J.F., Fox, N.A., 2004. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. Int. J. Psychophysiol. 53, 105–119.
Wong, P.C., Diehl, R.L., 2003. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. J. Speech Lang. Hear. Res. 46, 413–421.
Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422.
Xi, J., Zhang, L., Shu, H., Zhang, Y., Li, P., 2010. Categorical perception of lexical tones in Chinese revealed by mismatch negativity. Neuroscience 170, 223–231.
Xu, Y., 1997. Contextual tonal variations in Mandarin. J. Phon. 25, 61–83.
Xu, Y., Gandour, J.T., Francis, A., 2006a. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J. Acoust. Soc. Am. 120, 1063–1074.
Xu, Y., Krishnan, A., Gandour, J.T., 2006b. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601–1605.
Yip, M., 2003. Tone. Cambridge University Press, New York.
Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., Li, P., 2011. Cortical dynamics of acoustic and phonological processing in speech perception. PLoS One 6, e20963.
Zhang, L., Xi, J., Wu, H., Shu, H., Li, P., 2012. Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. Neuroreport 23, 35–39.
Zheng, H.-Y., Minett, J.W., Peng, G., Wang, W.S.-Y., 2012. The impact of tone systems on the categorical perception of lexical tones: an event-related potentials study. Lang. Cogn. Process. 27, 184–209.
Zheng, H.-Y., Peng, G., Chen, J.-Y., Zhang, C., Minett, J.W., Wang, W.S.-Y., 2014. The influence of tone inventory on ERP without focal attention: A cross-language study. Comput. Math. Methods Med. 2014, 1–8.