Journal of Neurolinguistics 33 (2015) 50–66

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

Second language experience modulates neural specialization for first language lexical tones

Benjamin D. Zinszer [a, b, *], Peiyao Chen [c], Han Wu [d], Hua Shu [d], Ping Li [a]

[a] Department of Psychology, Center for Brain, Behavior, and Cognition, Pennsylvania State University, USA
[b] Department of Brain and Cognitive Sciences, University of Rochester, USA
[c] Department of Communication Sciences and Disorders, Northwestern University, USA
[d] State Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University, China

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 16 September 2014
Accepted 18 September 2014
Available online 17 October 2014

Keywords:
Lexical tones
Categorical perception
Bilingualism
Chinese
Age of acquisition

ABSTRACT

Recent neuroimaging studies have revealed distinct functional roles of left and right temporal lobe structures in the processing of lexical tones in Chinese. In the present study, we ask whether knowledge of a second language (English) modulates this pattern of activation in the perception of tonal contrasts. Twenty-four native Chinese speakers were recruited from undergraduate and graduate students at Beijing Normal University, China. Participants listened to blocks of computationally manipulated /ba/ syllables which were varied to form within- and across-category deviants at equal acoustic intervals from a standard tone while their cortical blood oxygenation was measured by functional near-infrared spectroscopy (fNIRS). Blocks were analyzed for peak blood oxygenation (HbO) levels, and several linear models were estimated for these data, including effects of deviant tone type (within- or across-category), behavioral differences in tone identification, age of earliest exposure to English (spoken), and proficiency in English. Functional changes in HbO indicated a significantly greater response to within-category contrasts in right STG, consistent with previous findings. However, the effect of deviant type in left MTG was significantly modulated by the age of participants' earliest English exposure: Average across-category activation exceeded within-category activation only for participants exposed to English after 13 years of age. While previous research has established the importance of left MTG in the

* Corresponding author. Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA.
E-mail address: bzinszer@gmail.com (B.D. Zinszer).

http://dx.doi.org/10.1016/j.jneuroling.2014.09.005

categorical perception of lexical tones, our findings suggest that the functional specialization of this region is sensitive to second language experience, even in the processing of native language.

1. Introduction

Categorical perception of speech is one of the basic, defining features that distinguish language from other forms of animal communication (Hockett, 1963). Native-like language comprehension is marked by categorical perception (hereafter, CP) at multiple levels, such that continuous variation in a speech signal (e.g., place of articulation or voice onset time in consonants) is treated by language users as transitions between discrete, linguistically meaningful units or categories. Early research demonstrated that native English speakers judge continuous variation of stop consonants categorically (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), which is reflected as listeners' insensitivity to within-category variations but reliable discrimination of across-category variations of the same magnitude (two consonants on opposite sides of the categorical boundary).

1.1. CP and the functional hypothesis of tone processing

Suprasegmental features such as lexical intonation are also subject to these categorical perception effects (Gandour, 1983). Native speakers of Chinese, specifically, speakers of Mandarin in our study, perceive four linguistically salient pitch contours as distinct lexical tones. These contours are superimposed over an entire syllable and distinguish between minimal pairs that indicate meaning differences, just as phonemes do. Because of experiences with tonal categories, speakers of tonal languages such as Chinese and Thai, as compared with speakers of non-tonal languages, show an advantage at discriminating across-category contrasts like *dá* (Tone 2 – a rising pitch contour, meaning *arrive*) vs. *dà* (Tone 4 – a falling pitch contour, meaning *big*) in Mandarin Chinese but a disadvantage at discriminating within-category contrasts such as minor contour changes within the *dá* or *dà* categories (Xu, Gandour, & Francis, 2006).

Neurocognitive research in lexical tone processing has identified functional brain markers of CP. Native speakers of Thai (also a tonal language) and Chinese show lateralized mismatch negativity responses (MMN, a negative event-related potential at approximately 200 ms after the stimulus onset) to lexical tone contrasts from their native languages in an oddball paradigm (Kaan, Wayland, Bao, & Barkley, 2007; Xi, Zhang, Shu, Zhang, & Li, 2010; Yu, Wang, Li, & Li, 2014). Among native speakers of both Thai and Chinese, across-category contrasts elicit a left-lateralized MMN, and in native Chinese speakers, a greater right-lateralized MMN response for within-category contrasts (Xi et al., 2010). In a frequency domain analysis of electrophysiological responses, while non-native speakers showed greater alpha-band coherence in the right hemisphere when listening to Thai lexical tone contrasts, native Thai speakers showed this coherence bilaterally, suggesting differences in the involvement of left hemisphere for language-relevant vs. acoustic-only processing of the lexical tone contrasts (Kaan, Wayland, & Keil, 2013).

The lateralized response to acoustic vs. linguistic processing of lexical tones in native Chinese speakers has been further replicated and localized using fMRI. Zhang et al. (2011) showed that the hemodynamic response to across-category pairs of intonated syllables is greatest in the left middle temporal gyrus (MTG), consistent with left MTG's broader role in lexical retrieval, and the within-category contrast elicited a response in the right superior temporal gyrus (STG), suggesting acoustically based processing in the right hemisphere. Taken with the preceding electrophysiological evidence, this study presents a useful dissociative relationship for native-like CP of lexical tones: linguistically relevant contrasts (that is, discrimination between pitch contours across tone categories) elicits the greatest response in left hemisphere regions, specifically lexical-phonology processing areas such as MTG and STG, but linguistically irrelevant, purely acoustic, contrasts (that is, discrimination between pitch contours within the same tone category) elicits the greatest response in right hemisphere regions, specifically the primary auditory cortex in the STG. This distinction between left vs. right processing has been formulated by Gandour and his colleagues as the functional hypothesis of lexical tone processing (Gandour, 2006; Gandour et al., 2004). This left-right distinction, however, may not apply when considered in the context of the role of subcortical structures in tonal processing: Antoniou, Chandrasekaran, Deng, Li, and Wong (submitted for publication) recently showed that the auditory brainstem also plays a significant (and early) role in category perception of lexical tones.

1.2. Variation in categorical perception

Research in CP of lexical tones has focused on identifying commonalities among native speakers and comparing performance in identification or discrimination tasks between native and non-native speakers. This line of research aims at (1) understanding what behavioral patterns and neural markers define native-like CP and (2) exploring whether non-native speakers can attain native-likeness in CP through training. Most previous studies have so far regarded tone perception as relatively uniform among native speakers. However, this approach has ignored the potential individual variations in tone perception due to (a) effects of individual experience and (b) different learning environments.

In fact, native speakers show several sources of variation in CP of lexical tones. Native listeners identify different lexical tone contours with different degrees of CP, generally showing greater category sensitivity to rising (e.g., tone 2 of Mandarin) or falling contours (e.g., tone 4 of Mandarin) than to flat tones (e.g., tone 1 of Mandarin, Francis, Ciocca, & Ng, 2003; see also Abramson, 1979 regarding Thai tones). Such variations may be related to both acoustic and linguistic factors. Humans, in general, show greater acoustic sensitivity for changes in pitch over time than for relative pitch differences, and native speakers of a tonal language most often attend to these changes in pitch as category cues instead of absolute onset pitch, even though only the latter effectively distinguishes two flat tones (Francis et al., 2003).

A second variable underlying variation in native listeners' ability to identify lexical tones is linguistic context and adaptation to individual speakers. Several studies have shown that when native Cantonese speakers are asked to discriminate lexical tones from Cantonese, they perform significantly worse when provided with no sentential context (Francis et al., 2003; Ma, Ciocca, & Whitehill, 2006). Wong and Diehl (2003) also showed that native speakers were influenced by sentential context when identifying lexical tones and furthermore, that variation among native listeners' perception was related to the following: native speakers appear to develop an adaptive system for tone perception which is sensitive to differences among other speakers' productions, such as speakers with overall higher or lower pitch and wider or narrower bands of pitch variation in contoured tones. Native speakers can use these context clues from speakers to modify CP boundaries for lexical tones.

1.3. Bilingualism and categorical perception

Individual variation in tone perception may also arise as a result of language experience. Shifts in native language (L1) phonological categories have been widely documented for bilinguals immersed in a second language (L2) environment. English-French bilinguals and French-English bilinguals have both been observed to adjust L1 vowel and consonant categories towards L2 norms (Bullock & Gerfen, 2004; Flege, 1987). Phonological category shifts may arise as the resolution to competition between L1 and L2: for example, Major (1992) showed that Portuguese-English bilinguals displayed decrease in L1 phonological native-likeness proportional to their simultaneous gains in L2 native-likeness. The degree of cross-language shift may be modulated by age of L2 acquisition, as demonstrated by Italian–English bilinguals whose pre-voicing of /b/ changed towards English norms proportionally to their ages of English acquisition (MacKay, Flege, Piske, & Schirru, 2001). While these phonological category changes have been observed primarily in speech production, at least one study has also observed similar effects in speech comprehension. Korean-English bilinguals' performance in a Korean word recognition task under noisy conditions was linked to their relative English proficiency, such that greater L2 proficiency resulted in less successful L1 word recognition (Von Hapsburg & Bahng, 2009).

In addition to the cross-language L1–L2 interaction effects at the phonemic level, categorical shifts in L1 are also known to occur at the suprasegmental level. Prosodic patterns of native German and Dutch speakers show effects of bilingualism (De Leeuw, Mennen, & Scobbie, 2011; Mennen, 2004). In both studies, prosodic features (e.g., vowel length and prenuclear rise) shifted in L1 production towards the L2 norms such that prosody was no longer native-like in L2 or L1. These effects have previously been described as L1 attrition (e.g., De Leeuw et al., 2011; De Leeuw, Mennen, & Scobbie, 2012; De Leeuw, Schmid, & Mennen, 2010), due to the decreasing L1 native-likeness over time, but when these changes in L1 occur in concert with L2 change (as in the studies described above), there is reason to assume that cross-language transfer in fact shapes both L1 and L2 as bilinguals adapt to competing categorical norms of each language.

A recent study has further documented these transfer effects in an L1 environment where L1 attrition (specifically, loss of L1 not due to L2 influence) is an unlikely cause. Chang (2012) found that native speakers of English altered both segmental (voice onset times) and global (fundamental frequency) properties of English speech after only brief, intensive training in Korean as a second language. L2-to-L1 transfer in an L1 environment has also been observed in lexical semantics for Chinese college students studying English (Dong, Gui, & MacWhinney, 2005), raising the broader question of whether L1 language representations are, even in late adult learners, much more permeable to L2 influence than previously recognized.

1.4. Bilingualism and tone processing

In addition to the impact of bilingualism on phonological categories in general, research has also brought to light the role of first and second language experience in shaping lexical tone perception in particular. First, speakers of non-tonal languages, as compared with speakers of tonal languages, show a relative advantage in discriminating within-category tone contrasts, a hallmark of CP. While Chinese speakers showed greater sensitivity in detecting the across-category lexical tone contrasts in Chinese, native English speakers demonstrate greater sensitivity to within-category contrasts in Chinese (Xu et al., 2006).

Second, learners attend to different aspects of lexical tones based on their native language background. In a training study for Thai lexical tones, native speakers of Chinese tended to focus on pitch contour of Thai tones, but native speakers of English attended more on onset pitch difference (Kaan, Barkley, Bao, & Wayland, 2008). This tendency has also been observed in English-speaking learners of Mandarin Chinese, who rely on pitch onset differences to distinguish the Chinese tones (Chandrasekaran, Krishnan, & Gandour, 2007).

Third, individual differences and language experience also influence how people respond to training in lexical tones from a new language. In one study, English learners of Chinese who could eventually shift their focus from onset pitch towards pitch contour (or change in pitch) performed better in lexical tone training than English learners who perseverated in their focus on onset pitch (Chandrasekaran, Sampath, & Wong, 2010). In another study, after two days of training on the Thai lexical tones, native English speakers' EEG signals suggested greater effort in formation of new speech categories (relative to native Chinese speakers, as measured by power in the gamma band), but the English speakers also ultimately showed greater gains in signal coherence compared to the Chinese speakers, suggesting less effortful or more unified processing (Kaan et al., 2013).

The EEG study by Kaan et al. (2013) also suggests that learning new speech contrasts can alter brain responses even after a brief training period. During second language acquisition, new speech contrasts are assimilated into a learner's phonetic representations, and fMRI studies show that bilinguals accommodate these new representations by adapting the same brain regions already employed by their first language (Abutalebi & Green, 2007; Chee et al., 1999). Because native English and Chinese speakers differ in the way they attend to tonal information (and the significant differences in English and Chinese with regard to phonological categories), bilingual speakers of both languages may need to adapt to these competing cues (onset pitch vs. pitch contour). This difference could demand more accommodation, as opposed to simple assimilation, in the neural system than required of the processing of languages that are more similar to one another (see Li, 2014; Perfetti et al., 2007, for discussion). With regard to individual speakers, this accommodation may be realized as a strong effect of language experience on the processing of one's own native language. This change, if it occurs, may also be reflected in brain responses to CP of lexical tones. In the current study, we explore individual differences among native Mandarin Chinese speakers with regard to the neural correlates of CP for lexical tones, and we ask whether native Chinese speakers' experience with English as a second language is a significant predictor of these differences.

1.5. Functional near-infrared spectroscopy (fNIRS)

Previous research in this domain has relied mainly on behavioral testing, ERP, or fMRI methodologies. Functional near-infrared spectroscopy (fNIRS) is a neuroimaging technique that measures the hemodynamic response, similarly as functional magnetic resonance imaging (fMRI), but on the surface of the cerebral cortex. Like fMRI, fNIRS measures local changes in blood oxygen levels associated with the metabolic demands of firing neurons and the subsequent vascular response. While most biological tissue scatters light in the near-infrared spectrum, oxygenated and deoxygenated hemoglobin selectively absorb this light at two different wavelengths (approximately 695 and 830 nm). Therefore, relative changes in the amount of infrared light from a laser diode on the scalp can be measured to infer the relative changes in local blood oxygenation. This measurement is made by an optical detector placed approximately 3 cm away from a laser diode source (both on the scalp).

Although both fNIRS and fMRI measure a roughly analogous physiological response, they differ greatly in spatial resolution. A single fNIRS channel covers a region up to about 3 cm in diameter on the surface of the cortex, with the greatest signal being measured about halfway between the laser and detector and decreasing radially. Consequently, the spatial resolution of fNIRS is greatly reduced relative to fMRI (which measures up to a few millimeters of cortical regions). On the other hand, fNIRS offers a number of advantages; for example, it is a much less invasive measure (e.g., requiring no magnetic desensitization so patients with ferro-magnetic implants can be measured), is more robust against head motion (e.g., equipment is head mounted and recording can sometimes be done wirelessly), and is suitable for measuring special populations such as infants or patients who may need to be more mobile or who are claustrophobic (which would preclude them from being scanned under fMRI).

So far only one study has examined categorical perception of speech sounds using fNIRS. Minagawa-Kawai, Mori, Naoi, and Kojima (2007) and Minagawa-Kawai et al. (2011) tested Japanese-immersed infants for laterality of response to within- and across-category contrasts for vowel duration (a linguistically salient feature in Japanese speech). Their study also found a significant lateralization effect with greater activity in the left hemisphere for the across-category contrast. Examining the lateralization patterns in CP processing of Chinese lexical tones as in this study would lend further support to the feasibility and utility of fNIRS in future studies of categorical perception of speech.

1.6. The present study

In the present study, we examine the localization of lexical tone processing with native Mandarin Chinese speakers, similar to those participants sampled in previous lexical tone studies. Unlike previous studies, we account for variations in each participant's individual categorical perception boundaries. Most previous studies have focused on the processing of native speakers as if they all perform the CP processing to the same level or same degree. In our study, we attempt to understand the individual variations of processing even when all participants are native speakers, and to identify whether these variations are related to the participant's language history and experience with a second language, given the evidence we reviewed above.

Drawing on Minagawa-Kawai et al.'s (2007) paradigm for presentation of categorical perception stimuli with fNIRS measurement, we test within- and across-category contrasts of lexically intonated /ba/ syllables, using the same stimuli used by Xi et al. (2010) and Zhang et al. (2011; see details in Methods). We ask native speakers of Chinese to make judgments on eleven acoustically manipulated Chinese syllables that vary from rising tone (Tone 2) to falling tone (Tone 4) in equal intervals of pitch change. Participants passively listened to a stream of three versions of these /ba/ syllables while their oxygenated hemoglobin levels were measured using fNIRS. We compare the results of this experiment to the observed lateralization effect of Xi et al. (2010) based on ERP and the localized responses of Zhang et al. (2011) based on fMRI. We further examine whether the neural patterns in our data are due to the respective roles of English language proficiency and age of exposure to English as a second language, in particular, whether the change in blood oxygenation in the left MTG and the right STG will reflect the dissociation of functions observed in previous research, and whether individual variation in language history may modulate this dissociation.

2. Methods

2.1. Participants

Twenty-four native Mandarin Chinese speakers (18 female and 6 male) were recruited from Beijing Normal University in Beijing, China by advertisement on an electronic bulletin board (BBS). Mean participant age was 22.1 years (*SD* = 2.5), and all participants were screened to confirm right-handedness and minimal experience with any languages besides their native language and English as a second language. Table 1 describes participant demographics and language history. No participants reported neurological or learning disabilities, and they had normal or corrected-to-normal vision. The research was approved by the Institutional Review Board of Pennsylvania State University and the Research Ethics Committee of Beijing Normal University.

2.2. Materials

Participants completed a brief handedness scale in Chinese based on Snyder and Harris' (1993) English questionnaire, and this modified questionnaire consisted of nine items (e.g., "Brushing teeth"), with a handedness rating ranging from 1 (exclusively left-handed) to 5 (exclusively right-handed). Participants also completed a language history questionnaire (Li, Sepanski, & Zhao, 2006) in which they detailed their experiences with their first language, any additional languages, and any travel or educational experiences which may have significantly exposed them to other languages.

2.2.1. Acoustic stimuli

Acoustic stimuli from previous studies of Chinese lexical tone perception (e.g., Xi et al., 2010; Zhang et al., 2011) were used in the present study. Each stimulus is generated from a single /ba/ syllable recorded by a native Chinese speaker. This recorded syllable was computationally manipulated to fit eleven acoustically different pitch contours approximating the natural rising (tone 2) and falling (tone 4) tones in Chinese, with ten equally-spaced intermediate intervals. Using these stimuli, Xi et al. (2010) produced native-speaker norms for categorical perception of lexical tone. Contours 1 through 4 are consistently rated as rising tones (mean rating > 90%), and Contours 7 through 11 are consistently rated as falling tones (mean rating >90%). Contours 5 and 6 are considered boundary stimuli and may be categorized at random, which also varies individually by listeners.

Based on these previous categorical perception data, three of the intonated syllables were selected to form category contrasts: Contour 7 (a falling tone) was selected as the standard tone against which

Table 1
Participant demographics and tone identification task ratings.

Measure ($n = 23$)	Mean	S.D.
Age	22.1 y	2.5
English proficiency (self-rated)	3.5/7	0.78
Age of earliest English exposure	11.8 y	2.0
Contour 3 rating	0.05	0.13
Contour 7 rating	0.78	0.25
Contour 11 rating	0.97	0.07

Note. Excludes one participant who classified all stimuli as falling tones.

two types of deviant would be contrasted. The two deviant tones were then selected at equal acoustic distance from the standard Tone: Contour 3 (a rising tone) provided an across-category deviant and Contour 11 (another falling tone) provided a within-category deviant. Previous studies (Xi et al., 2010; Zhang et al., 2011) have confirmed that although the 7–3 and 7–11 stimulus pairs have equal acoustic distance between them, native Chinese speakers treat the 7–3 pair as across-category pair and the 7–11 pair as the within-category pair, and process these pairs differently (i.e., demonstrating the CP pattern) as evidenced by differences in the mismatch negativity response and localization of the hemodynamic response (see discussion in Section 1.1).

2.2.2. Syllable stream

A 26:40 (min:sec) stream was generated for every participant based on a stimulus presentation paradigm developed by Minagawa-Kawai et al. (2007) for an fNIRS study of categorical perception of Japanese vowels. This speech stream was composed of two sections contrasting either within- or across-category deviant syllables with standard syllables. The order of the within and across sections was counter-balanced between participants such that half of the participants heard the within section first and half heard the across section first.

Each section was composed of 20 blocks of syllables alternated with 20 blocks of silence (rest). All blocks were 20 s in duration and included 16 syllables (250 ms) with 1000 ms of silence between them. The 20 syllable blocks within a section alternated between standard blocks and deviant blocks. The contents of a deviant block were determined by the section: During the within-category section, deviant blocks contained a mixture of standard and within-category deviant syllables, and during the across-category section, deviant blocks contained a mixture of standard and across-category deviant syllables. In standard blocks, 16 repetitions of the standard syllable (Contour 7) were played. In deviant blocks, 8 standard syllables and 8 deviant syllables (Contour 3 or 11) were randomly ordered within the block.

2.3. Procedure

On the first day, all communication with participants was conducted through a native speaker of Chinese who was familiar with the experimental procedure and imaging methods. Before undergoing the informed consent, participants completed the handedness questionnaire to determine eligibility for the experiment. After consent was obtained, participants were seated in a comfortable chair in front of a personal computer.

Once seated, participants were familiarized with the NIRS device and the elastic cap containing the optodes that they would be wearing. The cap was placed on the participant's head directly above the eyebrows in the front and inion (IZ) in the back, which allows approximating the same overall placement of the optode array for each participant, covering temporal and parietal regions of the cerebrum. After the cap and optode array were secured to the participant's head, a Polhemus 3D digitizer was used to measure the participant's skull shape and the location of each optode on the scalp in 3D space. Five fiducial anatomical references were used according to the 10–20 system (Jasper, 1958): nasion (NZ), inion (IZ), the left and right preauricular points (A1 and A2), and CZ. Locations of each optode were also measured in the same coordinate system, allowing later inference of the approximate cortical location from which each measurement was taken.

When participants were oriented and comfortable in front of the computer, a set of clip-on earphones were attached over their ears. The participants were advised to relax and watch a brief movie while a stream of speech sounds were played in their earphones. The lights were dimmed and a silent nature film was played on the computer for the participants. At this time, functional NIRS measurement began with 20 s of baseline data collection. After the 20 s baseline, the speech stream commenced. fNIRS measurement continued until the end of the speech stream, for approximately 27 min fNIRS data were collected on a Hitachi ETG-4000 system using a bilateral array of sources and detectors. Source lasers emitted light at 695 and 830 nm, measured at a sampling rate of 10 Hz.

On the second day, participants returned to complete the behavioral tasks. The language history questionnaire was administered on the second day, followed by a two-alternative forced choice task to norm the original stimuli (as used in Xi et al., 2010). Participants listened to each of the eleven pitch-contoured syllables and indicated whether it was a rising or falling tone by pressing the up or down arrow, respectively, on a computer keyboard. Each syllable was played ten times for a total of 110 trials, randomly ordered.

2.4. fNIRS data processing

Raw optical density measures from the fNIRS device were converted to percent change in oxygenated hemoglobin (hereafter, HbO) according to the modified Beer–Lambert Law (Cope & Delpy, 1988) using the NIRS-SPM software package (Ye, Tak, Jang, Jung, & Jang, 2009). Time series data for each channel were high-pass filtered using a dynamic cosine transform (DCT) with a cut-off at 60 s (0.017 Hz) to remove drift and low-frequency noise. A 1 s moving average filter was used to smooth the time series and remove high frequency noise resulting from hardware and motion artifacts. Stimulus blocks were baselined to the average of the 5 s preceding block onset. Finally, hemodynamic responses were measured by extracting the maximum positive peak value in each block (as in Kovelman, Shalinsky, Berens, & Pettito, 2008).

The locations of each channel (the measurement occurring between a laser source and a detector) were estimated based on the averaged 3D model of ten participants' heads while wearing the NIRS cap. The averaged head dimensions and channel locations were entered into NIRS-SPM to project channel locations onto the cortical surface and standardize these locations to an MNI brain template using the NFRI function (Singh, Okamoto, Dan, Jurcak, & Dan, 2005). Fig. 3 depicts the estimated locations of 20 channels on the surface of the MNI brain. These locations were compared to an MNI atlas to estimate the relative coverage of cortical structures on each fNIRS channel (a region up to 3 cm in diameter, with greatest signal from the center of that region). Table 2 describes the primary anatomical structure underlying each channel and the percentage of that channel's coverage area that the anatomical structure occupies (Fig. 1).

Twenty channels were examined for their coverage of the regions of interest (ROI), specifically left MTG and right STG based on the previous studies of lexical tone processing (Zhang et al., 2011). The best two channels were selected for each ROI based on their coverage of one region (either MTG or STG) and exclusion of the other.

Table 2
Coverage of anatomical structures by NIRS channels (MNI atlas).

Channel	Centroid (MNI)			BA	Structure	% Coverage
1	−64	−53	−16	37	FG	70%
2	**−71**	**−24**	**−12**	**21**	**MTG**	**83%**
3	**−63**	**2**	**−11**	**21**	**MTG**	**76%**
6	−69	−44	−1	21	MTG	46%
7	−69	−15	3	22	STG	59%
10	−64	−58	11	37	FG	59%
11	−69	−30	17	22	STG	69%
12	−65	0	21	43	SCA	60%
15	−65	−49	29	22	STG	44%
16	−67	−20	34	2	S1	58%
24	67	−1	−7	21	MTG	81%
25	72	−27	−9	21	MTG	96%
26	65	−55	−14	37	FG	91%
29	**70**	**−11**	**8**	**22**	**STG**	**78%**
30	71	−41	4	22	STG	57%
33	67	1	21	43	SCA	63%
34	**71**	**−31**	**19**	**22**	**STG**	**64%**
35	64	−60	13	37	FG	55%
38	68	−13	34	43	SCA	49%
39	67	−43	33	40	SMG	69%

Note. FG: fusiform gyrus; MTG: middle temporal gyrus; STG: superior temporal gyrus; SCA: subcentral area; S1: primary somatosensory cortex; SMG: supramarginal gyrus. Bold entries indicate channels selected for analysis in the present study.

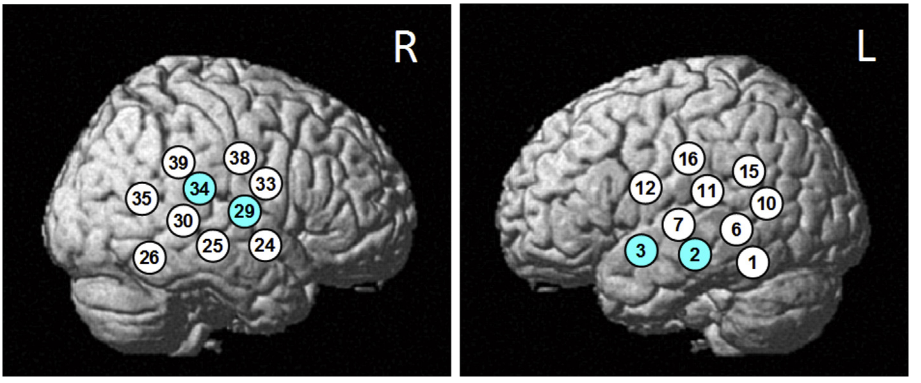

Fig. 1. Plot of channel locations on MNI brain template. Twenty channels (from a 44 channel array) were selected to represent left and right temporo-parietal regions. Two channels were further selected for measurement of left middle temporal gyurs (MTG) and right superior temporal gyrus (STG), highlighted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Categorical perception of lexical tones

Each participant's responses in the 2-alternative forced-choice pitch contour identification task were tallied across ten repetitions per contour, producing an individual categorical perception curve. One participant was excluded at this stage for rating all stimuli as falling tones greater than 50% of the time. Mean responses across the remaining 23 participants (with 95% C.I.) are depicted in Fig. 2A.

Four participants deviated from typical native-like category boundaries by classifying Contour 7 (a falling tone) as rising in 50% or more of trials, as illustrated in Fig. 2B. Given the lack of sentential context and Contour 7's proximity to relatively level contours (4–6), these idiosyncratic variations in category boundaries may reasonably fall within the range of native speakers' perceptual variability. Nonetheless, previous experiments using an identical paradigm have shown high confidence among native speakers for the categories of these specific stimuli. The group-level agreement for Contour 7 was significantly lower than what Xi et al. (2010) found for a similar population of native Chinese speakers ($t(22) = 2.15$, $p = 0.042$ for H_0: $\mu = 0.90$). Some exploration of this discrepancy is presented in the Discussion, but for the purpose of this experiment, we accept a mean rating of 0.78 ($SD = 0.25$) for Contour 7 as sufficient for establishing its category membership. Furthermore, we included individuals' unique ratings of Contours 3, 7, and 11 as predictors (interacting with deviant type) in the models of the fNIRS data to account for effects that perceptual variation among native Chinese speakers may have on category-specific processing.

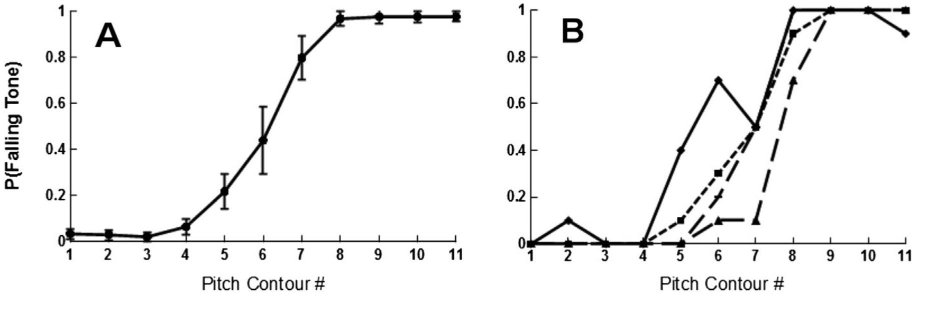

Fig. 2. (A) Mean responses in categorical perception task for eleven pitch-contoured syllables (error bars represent 95% CI). (B) Individual curves for four participants who differed from native-like category boundary.

Table 3
Correlations between English language history and tone judgments.

Language history variable	T3	T7	T11
English proficiency (self-rated)	−0.07	−0.11	0.15
Age of earliest english exposure	−0.20	0.14	**0.45/0

Note. **$p < 0.05$, $R = 0$ after removing outlier. All other $p > 0.30$.

Given the inter-participant variation in performance on the category identification task, we also tested whether category judgments for any of the three critical tones were correlated with second language experience. Table 3 outlines the Pearson R correlations, of which the only significant correlation occurred between Contour 11 (T11) and Age of Earliest English Exposure (AOEE). This relationship was driven entirely by a single participant who rated T11 as a falling tone only 70% of the time (as compared to 90–100% in other participants). Consequently, no strong connection was found between behavioral performance in tone identification and history of English language learning.

3.2. Neural correlates of categorical perception

Previous studies have found significant main effects for functional lateralization in native speakers of Chinese based on the mismatch negativity (Xi et al., 2010) and localization of the BOLD signal (Zhang et al., 2011). In the present analyses, we compare our results to these two accounts of functional dissociation. First we apply analogous tests to the fNIRS data, comparing hemispheres and specific ROIs for the main effects of the deviant types (across-vs. within-category blocks). Next we allow the deviant type to interact with variables describing participants' individual L2 language learning histories. Previous research has identified age of L2 acquisition (MacKay et al., 2001) and L2 proficiency (Major, 1992; Von Hapsburg & Bahng, 2009) as important predictors of L1 change in bilinguals due to L2 influence. We use age of earliest exposure to spoken English (AOEE) and self-rated proficiency ratings for speaking and listening to English (L2 Proficiency) as measures of individual participants' language experience.

3.2.1. Hemisphere comparison

We conducted a repeated-measures ANOVA to test the lateralization effect observed by Xi et al. (2010). Participant ratings for Contours 3, 7, and 11 and counter-balancing condition were entered as controls (interacting with deviant type) to account for individual differences in the category boundary. No significant main effect was observed for deviant type ($F(1,9161) = 1.466$, $p = 0.226$) or Hemisphere ($F(1,9161) = 1.617$, $p = 0.204$). Deviant type and Hemisphere also showed no interaction ($F(1,9161) = 1.953$, $p = 0.162$). Fig. 3A shows the mean peak oxygenated hemoglobin (HbO) across participants for each deviant type. The results of this ANOVA failed to replicate the lateralization patterns observed in Xi et al. (2010) for the mismatch negativity.

The participant ratings of Contours 3, 7, and 11 (entered as controls for inter-participant differences in category identification) were all highly significant predictors. Main effects for Contours 7 and 11 significantly predicted overall variation in peak oxygenation (T7: $F(1,9161) = 27.418$, $p < 0.001$; T11: $F(1,9161) = 438.796$, $p < 0.001$), and Contours 3 and 11 significantly interacted with deviant type (T3: $F(1,9161) = 11.210$, $p < 0.001$; T11: $F(1,9161) = 29.711$, $p < 0.001$). The directions of these effects indicated that overall activity decreased with higher ratings for T11 (more strongly categorical) and differences in activity due to T7 ratings were negligible (0.0005% change HbO for every 10% change in the tone identification, though statistically significant). Higher ratings for T3 (less strongly categorical) were associated with a decrease in activity for the within-category deviant condition but not across-category, and higher T11 ratings (again, more strongly categorical) were associated with a greater decrease in activity for the across-category than the within-category deviant.

3.2.2. ROI comparison

We also conducted a repeated-measures ANOVA to test the dissociation of function observed for left middle temporal gyrus (MTG) and right superior temporal gyrus (STG), as identified by Zhang et al.

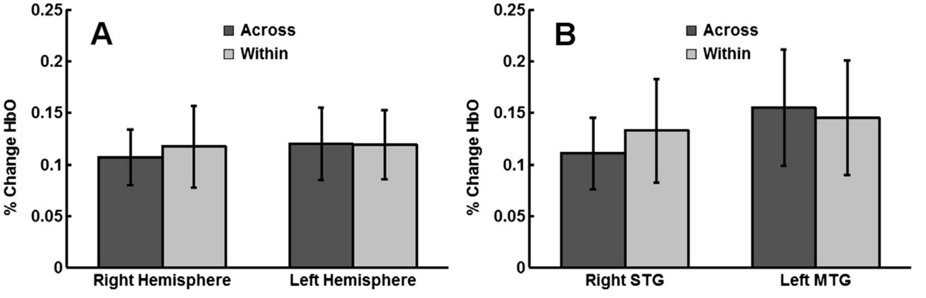

Fig. 3. Mean peak HbO responses for each deviant type in (A) left and right hemispheres and (B) left middle temporal gyrus (MTG) and right superior temporal gyrus (STG). Error bars represent 95% confidence intervals of the mean.

(2011). Participant ratings for Contours 3, 7, and 11 and counter-balancing condition were again entered as controls (interacting with deviant type). No significant main effect was observed for deviant type ($F(1,1817) = 0.001$, $p = 0.978$) or ROI ($F(1,1817) = 2.251$, $p = 0.134$). Deviant type and ROI also showed no interaction ($F(1,1817) = 0.569$, $p = 0.451$). Fig. 3B shows the mean peak HbO across participants for each deviant type. The results of this ANOVA also failed to replicate the dissociation between left MTG and right STG.

We again examined the effects of the controls for participant ratings of the critical contours, which were highly significant in the hemisphere-level model. In these ROIs, only T7 and T11 were significant predictors. T7 had a statistically significant ($F(1,1817) = 6.029$, $p = 0.014$) but negligible main effect (−0.0005% per 10% change in rating). As in the hemisphere comparison, higher T11 ratings were associated with lower activity overall, but this decrease was greatest in the across deviant (T11: $F(1,1817) = 92.850$, $p < 0.001$; T11 × deviant type: $F(1,1817) = 8.123$, $p < 0.01$).

At this stage of the analysis, neither the ERP or fMRI results have been replicated in our data. Although the individual participants' ratings of the three critical tone contours were significant predictors of variation in HbO levels, accounting for these differences did not result in the significant effects of deviant type predicted by previous research. More specifically, Contour 11 ratings (a falling contour) were a consistently strong predictor of functional activity such that more strongly categorical behavior (higher rating for T11) resulted in less overall functional activity, with a significantly greater decrease in activity for the across-category deviant.

3.2.3. Second language predictors

Given the preceding null results from our fNIRS data with respect to localization of CP and given the previous studies of individual variations in categorical perception and the goals of our study (see Sections 1.2–1.4), we entered the two second language predictors (AOEE and L2 Proficiency) into the repeated-measures ANOVA, interacting with ROI and deviant type. Age of earliest exposure to spoken English (AOEE) had a highly significant main effect ($F(1,1809) = 49.690$, $p < 0.001$) as well as interactions with ROI ($F(1,1809) = 4.338$, $p = 0.037$) and deviant type ($F(1,1809) = 6.738$, $p = 0.010$). There was no significant three-way interaction between AOEE, ROI, and deviant type ($F(1,1809) = 1.518$, $p = 0.218$).

L2 Proficiency as measured by self-rated proficiency in English (speaking and listening) also had a significant main effect ($F(1,1809) = 22.387$, $p < 0.001$) but failed to interact with ROI ($F(1,1809) < 0.001$, $p = 0.988$) or deviant type ($F(1,1809) = 0.170$, $p = 0.680$). However, because AOEE and L2 Proficiency are moderately correlated across participants ($R = -0.40$, $p = 0.058$), we retained the L2 Proficiency terms in the model so that estimated effects of AOEE would represent a unique predictor apart from any effects of L2 Proficiency.

The T7 and T11 control variables were again also significant predictors, following the same pattern as observed in the Hemisphere and ROI analyses (which did not include the second language predictors). The main effect of T7 was significant ($F(1,1809) = 6.285$, $p < 0.05$, −0.003% per 10% rating change), and T11 had both significant main effects ($F(1,1809) = 96.801$, $p < 0.001$) and interacted with deviant type ($F(1,1809) = 8.453$, $p < 0.01$) such that higher T11 ratings were again with less activity only for the across-category deviant.

3.3. Language history model

A linear regression model was generated based on the variables explored in the Second Language Predictors test. Due to the significant influence of the control variables (including L2 Proficiency), these variables were retained in the model, and the significant AOEE × ROI and AOEE × deviant type interactions were included. The values for the contour ratings were set for perfect categorical perception ($T3 = 0.0$; $T7 = 1.0$; $T11 = 1.0$) to control for functional correlates of performance in the behavioral task. This model accounted for 50% of the variance in the observed data, which allowed us to generate predictions about the relationship between AOEE and the relative values of the deviant types in each ROI. Fig. 4 illustrates the model's prediction for across- and within-category deviants at left MTG and right STG for various values of AOEE.

Predicted responses were generated at the lowest observed value of L2 Proficiency (1.5 out of 7) to represent minimal English knowledge while still varying AOEE. In the model's predicted responses, right STG showed greater response to within- than across-category contrasts overall ($M = 0.088$), and within the range of observed data (AOEE 7–15 y), the within-category response remains greater. By contrast, the relative magnitudes of across- and within-category response were reversed in the left MTG. For early ages of English onset (AOEE ≤ 13), within-category responses exceeded across-category responses. However for later AOEE, response to across exceeded within, reflecting the pattern of dissociation that characterizes native, monolingual speakers. These patterns, specifically the impact that AOEE has on within- vs. across-category responses, suggest that native phonological or tonal processing is permeable, and neural responses as measured here by fNIRS can capture such permeability with a high degree of accuracy.

4. Discussion

4.1. Summary

The present study is a first attempt to use the functional near-infrared spectroscopy (fNIRS) to investigate categorical perception of lexical tones. We specifically revisited two previous investigations of categorical perception for Mandarin Chinese lexical tones that have relied on the ERP (Xi et al., 2010) and fMRI (Zhang et al., 2011) methodologies. These studies, along with other research in this field, have indicated dissociable language-specific processing of lexical tones in native Chinese speakers based on across- and within-category contrasts. Using the same stimuli employed in these studies, we reproduce similar group-

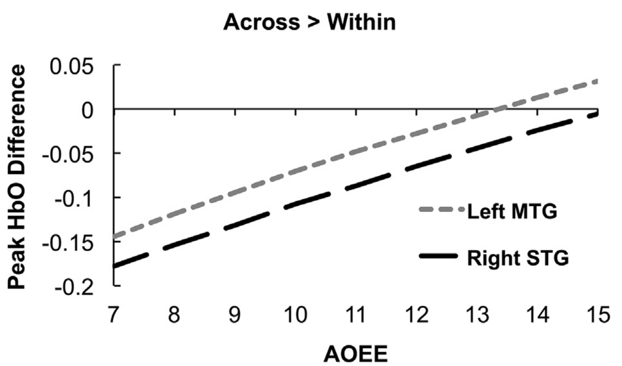

Fig. 4. Predicted values for difference in peak HbO response in each ROI based on the linear model including participants' language history. The across > within contrast in left MTG crosses zero around an AOEE (age of earliest English exposure) of 13 years old. Right STG remains below zero (Within > Across) across all AOEE.

level behavioral results for categorical perception. At the same time, we have also identified significant individual-level variation in the perception of these tones. Further, we find much variation in the hemodynamic response to these stimuli, in which possible lateralization and ROI results are obscured by individual differences, even after controlling for performance in the behavioral perception task.

To reconcile our findings with previous studies, we develop an account of individual differences in lexical tone processing by taking into account the learner's language history, in particular, experience with English as a second language. We demonstrate that previously observed neural patterns of categorical perception can vary as a function of both age of onset (AOEE) and proficiency of English, which is reflected as different hemodynamic responses to across-category tone contrasts. Our findings suggest that knowledge of a second language (in this case English) may affect the processing patterns in one's own language.

4.2. Individual variation in CP

The group-level tone identification curve in our study generally resembles a typical example of the categorical perception curves for lexical tones found in many previous studies, but the exact location of the category boundary significantly differs from Xi et al.'s (2010) previous study using the same stimuli and sampling from the same population (college and graduate students in Beijing). Further, we see considerable individual variation in both the location and slope of the category boundary between the two lexical tones in this task.

These variations in categorical perception were highly significant predictors of differences in functional activity (peak HbO levels) between participants in each analysis reported in this study, particularly for Contour 11. Generally, this finding concurs with previous studies that have found shifts in the electrophysiological response of learners as they become more native-like in discrimination of L2 lexical tones (Kaan et al., 2007; Kaan, Barkley, Bao, & Wayland, 2008). The present study, however, links these differences to normal variation among native speakers of Chinese rather than longitudinal development.

The previous studies of Chinese lexical tone CP in native speakers do not report on individual-level performance and thus tend to overlook variation among native speakers in the shape of identification curves that define CP boundaries and any possible neural correlates of this variation. Xi et al. (2010) find an abrupt categorical boundary between rising and falling tones, but there is no variance of the behavioral norming sample reported for each contour. Zhang et al. (2011) also report significant differences between within- (Contours 7 vs. 11) and across-category (Contours 7 vs. 3) discrimination in an AX discrimination task (using the same stimuli as Xi et al., 2010), although the considerable variation of within-category discrimination scores are not further explored.

Although our behavioral task identified four participants with relatively atypical category boundaries based on identifying Contour 7 as a rising tone in half or more of the trials, it was not the Contour 7 rating that most strongly predicted overall variation in neural response. The Contour 11 rating was a consistent behavioral predictor of functional activity and significantly interacted with deviant-type across all analyses, linking performance in an L1 category identification task to the neural signatures of categorical perception.

In effect, we find that participants who more confidently (100% vs. 90%) rated Contour 11 as a falling tone show overall less functional response to the tone stimuli in general, particularly for the across-category deviants. The inverse relationship between behavioral performance and functional activity is often attributed to neural efficiency (or a reduction in effort for processing), as in Kaan et al.'s (2013) observation that successful learners showed decreases in gamma-band electrophysiological activity and an increase in signal coherence. The reliability of the behavioral predictors' effects on functional activity in the present study suggests that native speakers and learners may not greatly differ in this respect—individual differences in native categorical perception behavior are linked to individual variation in processing effort.

4.3. Second language effects on CP

Given the role of English language learning observed in this study, we considered that behavioral differences in CP between our study and previous studies of native Chinese speakers might arise from a more rigorous screening process in previous studies whereby participants with even modest second language experience may have been excluded. However, English language history variables (AOEE and Proficiency) failed to correlate with the (behavioral) variations in the category identification ratings of the three critical tone contours. Our analyses included the significant effects of both the behavioral differences in CP and the differences in L2 experience, demonstrating that L2 variables remain significant predictors even after accounting for individual behavioral differences and confirming that these two predictors are minimally related, as the estimated effects of the behavioral predictors on functional activity remained relatively unchanged with the inclusion of the L2 predictors.

After accounting for variables describing individual differences and second language experience, we are indeed able to replicate the dissociated across- and within-category effects in the ROIs identified by Zhang et al. (2011), even though the pattern is somewhat attenuated. This replication shows that we can use fNIRS in future studies of lexical tone CP (e.g., see Minagawa-Kawai et al., 2007 discussed in the Introduction). Reproducing the dissociation observed by Zhang and colleagues supports their account that left MTG underlies the language-specific processing of lexical tones, as defined by the linguistic boundary between tone categories, but our additional finding that this response is modulated by experience with English as an L2 provides unique convergent evidence on the linguistic role of the left MTG. That is, left MTG's response to lexical tone categories depends on an individual's specific language learning experience, even across a sample of native Chinese speakers with little to moderate English proficiency. Crucially, these results demonstrate that differences in L1 processing may arise independently of changes in L1 behavior, as neural systems supporting L1 are re-shaped to accommodate new input (e.g., a second language) while behavioral performance continues to maintain native L1 proficiency.

More broadly, the current study provides further evidence that second language phonetic representations influence the perception of the first language. Mounting evidence from recent studies is showing L1–L2 interactions at the lexical, semantic, and syntactic levels (see Grosjean & Li, 2013; Li, 2014; Dussias & Sagarra, 2007), and that one's native language (L1) is much more permeable (i.e., subject to change) than previously thought. These interaction or competition effects are particularly known in the domains of lexical semantics where L1 lexical semantics have been shown to change as a function of L2 experience (see Dong et al., 2005; Zinszer, Malt, Ameel, & Li, 2014). L2-to-L1 phonological or prosodic transfer effects have also been shown to occur in circumstances of L2 immersion (e.g., Bullock & Gerfen, 2004; Flege, 1987) and profound L1 attrition (e.g., De Leeuw et al., 2012; Mennen, 2004). However, prior to this study there has been little evidence of L2-to-L1 influence occurring as the direct result of L2 learning without immersion (Chang, 2012 being the notable exception), especially at the suprasegmental level (i.e., tonal processing). There is now good reason to believe that segmental and suprasegmental phonological categories are subject to similar influences of age of L2 onset (MacKay et al., 2001) and proficiency (Major, 1992; Von Hapsburg & Bahng, 2009), as demonstrated for lexical tones in the present study.

The specific circumstances of our findings are particularly interesting for the degree of sensitivity that L1 neural representations may show to L2 influence. L1 phonological processing capacities, even at a late stage of adult representation, are not static but remain permeable, subject to impacts from one's experience with the L2. This influence is exerted well after the typical age after which infants learn native-like lexical tone discrimination (9 months old; Mattock, Molnar, Polka, & Burnham, 2008). Plasticity for these lexical tone representations seems to be maintained throughout childhood, however, in contrast to hypothesis based on L1 speech perception research (see Kuhl, 2004 for review). This is also consistent with recent findings from the neuroplasticity of bilingualism, where there is mounting evidence that not only functional, but also anatomical changes are occurring in the learning brain across the lifespan (see Li, Legault, & Litcofsky, 2014, for a review). In the present study, only participants with the latest ages of English onset (>13 years old) showed native-like left MTG activity. Processing differences in categorical perception of tones have also been previously observed in Chinese-speaking children who are at risk for dyslexia, who show a reduced left-lateralized MMN in response to the same lexical tone contrasts (Zhang et al., 2012). Both the adults in our study and the children in Zhang et al.'s (2012) study share a unique language experience relative to other native speakers, reflected in their processing of lexical tones and their neural adaptations for categorical perception.

Finally, none of the participants in the present study have been immersed in an English-speaking environment beyond their English classrooms in China. This is surprising, in light of the finding that language immersion provides a critical context for development of native-like representation and processing (see Kuhl, 2010 for review). The degree of cross-language influence appears to be more age-related than proficiency-related in our study, given that age of English onset served as a better predictor of left MTG response than English proficiency (but see Hernandez & Li, 2007; Li & Tokowicz, 2012 for discussion of the relative role of age of acquisition vs. proficiency in neuroimaging findings). Given that the current sample included only participants with relatively little English experience, age of acquisition may simply be the dimension along which participants showed the greatest variation. In future research it is necessary to capture a broader range of Chinese-English bilinguals at varying ages of English onset, proficiency, and immersion experiences, to determine the relative importance of each factor in the perception of lexical tones and the neural processes supporting native-like perception. With the ubiquity of bilingualism, future studies should also carefully delineate how native speaker status is defined, as many native speakers of a language may have experience with an L2 that influences neural representations in *both* languages. This is especially important for cultures where second language education (or even advanced bilingualism) is the norm, as native speakers may be best represented by a cross-section of bilinguals and monolinguals with varied language histories.

Acknowledgments

We wish to thank Thomas Holt and Shuang Song for assistance with data collection and coding, as well as Linjun Zhang for comments on a previous draft. This research was supported by the East Asia and Pacific Summer Institute grant from the National Science Foundation (#OISE-1210179) to BZ, a grant from the Confucius Institute at Penn State University to BZ, and a grant from the National Science Foundation (#BCS-1057855) to PL.

References

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom, & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 127–134). London: Academic Press.

Abutalebi, J., & Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242–275.

Antoniou, M., Chandrasekaran, B., Deng, Z., Li, P., & Wong, P. C. M. Brainstem responses to speech are category selective (submitted for publication).

Bullock, B. E., & Gerfen, C. (2004). Phonological convergence in a contracting language variety. *Bilingualism: Language and Cognition, 7*(2), 95–104. http://dx.doi.org/10.1017/S1366728904001452.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148–156. http://dx.doi.org/10.1016/j.brainres.2006.10.064.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456–465. http://dx.doi.org/10.1121/1.3445785.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268. http://dx.doi.org/10.1016/j.wocn.2011.10.007.

Chee, M. W., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W., Thiel, T., et al. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron, 23*(1), 127–137.

Cope, M., & Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical and Biological Engineering and Computing, 26*(3), 289–294.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2011). Singing a different tune in your native language: first language attrition of prosody. *International Journal of Bilingualism, 16*(1), 101–116. http://dx.doi.org/10.1177/1367006911405576.

De Leeuw, E., Mennen, I., & Scobbie, J. (2012). Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism, 17*(6), 683–700. http://dx.doi.org/10.1177/1367006912454620.

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition, 13*(1), 33–40. http://dx.doi.org/10.1017/S1366728909990289.

Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition, 8*(3), 221–238. http://dx.doi.org/10.1017/S1366728905002270.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition, 10*(1), 101. http://dx.doi.org/10.1017/S1366728906002847.

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 47–65.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029–1044.

Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics, 11*(2), 149–175.

Gandour, J. (2006). Brain mapping of Chinese speech prosody. In P. Li, L. Tan, E. Bates, & O. Tzeng (Eds.), *Handbook of east Asian psycholinguistics* (Vol. 1, pp. 308–319). Cambridge, UK: Cambridge University Press.
Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23*, 344–357.
Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism*. John Wiley & Sons.
Hernandez, A. E., & Li, P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650. http://dx.doi.org/10.1037/0033-2909.133.4.638.
Hockett, C. F. (1963). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of language* (pp. 1–22). Cambridge, MA: MIT Press.
Jasper, H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10*(2), 370–375.
Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neuroscience, 9*, 53. http://dx.doi.org/10.1186/1471-2202-9-53.
Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: an event-related potential study. *Brain Research, 1148*, 113–122. http://dx.doi.org/10.1016/j.brainres.2007.02.019.
Kaan, E., Wayland, R., & Keil, A. (2013). Changes in oscillatory brain networks after lexical tone training. *Brain Sciences, 3*(2), 757–780. http://dx.doi.org/10.3390/brainsci3020757.
Kovelman, I., Shalinsky, M. H., Berens, M. S., & Petitto, L. (2008). Shining new light on the brain's “bilingual signature”: a functional near infrared spectroscopy investigation of semantic processing. *NeuroImage, 39*(3), 1457–1471.
Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews: Neuroscience, 5*(11), 831–843. http://dx.doi.org/10.1038/nrn1533.
Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713–727. http://dx.doi.org/10.1016/j.neuron.2010.08.038.
Li, P. (2014). Bilingualism as a dynamic process. In B. MacWhinney, & W. O'Grady (Eds.), *Handbook of language emergence*. John Wiley & Sons, Inc.
Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex, 58*, 301–324. http://dx.doi.org/10.1016/j.cortex.2014.05.001.
Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: a web-based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210.
Li, P., & Tokowicz, N. (2012). The psycholinguistics of second language acquisition. *Routledge encyclopedia of second language acquisition*. New York, NY: Taylor & Francis/Routledge.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431–461.
Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America, 120*(6), 3978–3987. http://dx.doi.org/10.1121/1.2363927.
MacKay, I. R., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America, 110*(1), 516–528. http://dx.doi.org/10.1121/1.1377287.
Major, R. (1992). Losing English as a first language. *The Modern Language Journal, 76*(2), 190–208. http://dx.doi.org/10.1111/j.1540-4781.1992.tb01100.x/abstract.
Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367–1381. http://dx.doi.org/10.1016/j.cognition.2007.07.002.
Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32*(4), 543–563. http://dx.doi.org/10.1016/j.wocn.2004.02.002.
Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *The Journal of Neuroscience, 27*(2), 315–321. http://dx.doi.org/10.1523/JNEUROSCI.1984-06.2007.
Minagawa-Kawai, Y., Van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex, 21*(2), 254–261. http://dx.doi.org/10.1093/cercor/bhq082.
Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.-H. (2007). Reading in two writing systems: accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition, 10*(2), 131. http://dx.doi.org/10.1017/S1366728907002891.
Singh, A. K., Okamoto, M., Dan, H., Jurcak, V., & Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage, 27*(4), 842–851. http://dx.doi.org/10.1016/j.neuroimage.2005.05.019.
Snyder, P. J., & Harris, L. J. (1993). Handedness, sex, familial sinistrality effects on spatial tasks. *Cortex, 29*(1), 115–134. http://dx.doi.org/10.1016/S0010-9452(13)80216-X.
Von Hapsburg, D., & Bahng, J. (2009). Effects of noise on bilingual listeners' first language (L1) speech perception. *SIG 6 Perspectives on Hearing and Hearing Disorders: Research and Diagnostics, 13*, 21–26. http://dx.doi.org/10.1044/hhd13.1.21.
Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46*(2), 413–421.
Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231. http://dx.doi.org/10.1016/j.neuroscience.2010.06.077.
Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063. http://dx.doi.org/10.1121/1.2213572.
Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *NeuroImage, 44*(2), 428–447. http://dx.doi.org/10.1016/j.neuroimage.2008.08.036.

Yu, K., Wang, R., Li, L., & Li, P. (2014). Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Frontiers in Human Neuroscience*. Article #729, http://dx.doi.org/10.3389/fnhum.2014.00729.

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*(6), e20963. http://dx.doi.org/10.1371/journal.pone.0020963.

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., et al. (2012). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry, 53*(8), 874–882.

Zinszer, B. D., Malt, B. C., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in Psychology, 5*(01203). http://dx.doi.org/10.3389/fpsyg.2014.01203.