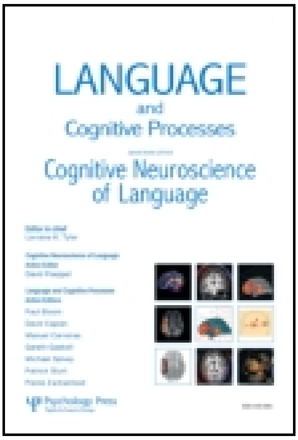

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study

Hong-Ying Zheng [a], James W. Minett [a], Gang Peng [a b] & William S-Y. Wang [a]

[a] Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

[b] Department of Linguistics and State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Pokfulam, Hong Kong

Version of record first published: 24 May 2011

LANGUAGE AND COGNITIVE PROCESSES
2012, 27 (2), 184–209

Psychology Press
Taylor & Francis Group

The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study

Hong-Ying Zheng[1], James W. Minett[1], Gang Peng[1,2], and William S-Y. Wang[1]

[1]Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong
[2]Department of Linguistics and State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Pokfulam, Hong Kong

This study investigates the categorical perception (CP) of pitch contours (level and rising) by native listeners of two tone languages, Mandarin and Cantonese, for both speech and nonspeech. Language background was found to modulate participants' behavioural and electrophysiological responses to stimuli presented in an active oddball paradigm, comprising a standard and two equally spaced deviants (within- and across-category). The stimuli were divided into two sets according to the results of a two-alternative forced-choice identification test: a *rising set*, using a standard that listeners identified as high rising tone, and a *level set*, using a standard that listeners identified as high level tone. For the rising set, both groups of listeners exhibited CP in terms of their behavioural response. However, only Cantonese listeners exhibited a significant CP effect in terms of P300 amplitude. For the level set, the behavioural data revealed a shift in category boundary due, in part, to the range–frequency effect. According to the d' scores, the CP effect elicited from Mandarin listeners was greater for nonspeech stimuli than for speech, suggesting the presence of a

Correspondence should be addressed to James W. Minett, Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: jminett@ee.cuhk.edu.hk

The work described in this paper was partially supported by research grants awarded to William S-Y. Wang by the Shun Hing Institute of Advanced Engineering of The Chinese University of Hong Kong (Project No. BME-8115020) and by the Research Grants Council of the Hong Kong Special Administrative Region, China.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2010.520493

psychophysical boundary. There was no such behavioural contrast for Cantonese listeners. However, Cantonese listeners exhibited a significant CP effect in P300 amplitude that was influenced by the range–frequency effect, as well as a possible secondary phonological boundary. P300 amplitude is believed to index the ease of discrimination of speech stimuli by phonological information. We conclude that Cantonese listeners engaged phonological processing in order to discriminate speech stimuli more efficiently than Mandarin listeners. These findings may be due to the different tonal inventories of Mandarin and Cantonese, with Cantonese listeners required to make finer distinctions in perception of pitch height and slope than Mandarin listeners in order to discriminate the denser tone system of Cantonese.

Keywords: Categorical perception; Lexical tone; Chinese Mandarin; Cantonese; Event-related potentials; P300; Oddball paradigm.

INTRODUCTION

Speech sounds vary significantly across different contexts and speakers, yet listeners typically perceive speech without difficulty. Categorical perception (CP) of speech occurs when listeners map speech sounds to discrete categories (Liberman, Harris, Hoffman, & Griffith, 1957), resulting in better discrimination of stimuli across category boundaries than equivalently separated stimuli within the same category. CP of speech develops from early infancy (Dehaene-Lambertz & Gliga, 2004). Newborns are able to discriminate most sound contrasts that are used categorically in the world's languages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). However, during the first year of life, as a result of exposure to the native language, discrimination of most non-native contrasts is lost (Gervain & Mehler, 2010; Mattock & Burnham, 2006; Werker & Tees, 1984), whereas native contrasts are maintained or even enhanced (Burnham, Earnshaw, & Clark, 1991; Kuhl et al., 2006). Such developmental reorganisation of perception during infancy—to favour native phonemic contrasts—is believed to be due largely to the use of statistical learning to track the frequencies of sound tokens in the input (Best & McRoberts, 2003; Kuhl, 2004).

Studies of Japanese infant perception of phonemic vowel length have pointed to the neural reorganisation of the infant brain (Minagawa-Kawai, Mori, Naoi, & Kojima, 2007). CP of phonemic difference develops from a generic auditory circuit at 6–7 months, but switches over to a more linguistic circuit after 12 months (Minagawa-Kawai et al., 2007). A similar switch from sensory circuits to linguistic circuits has been found for infant colour CP (Franklin et al., 2008). The acquisition of speech CP is based not only on analyses of the distributions of speech sounds, but also on functional analyses of the categories of those sounds. For example, in Russian, voice-onset-time (VOT) features contrast the phonemic categories /t/ and /d/. In other languages, such as Korean, these same features provide no phonemic contrast, thus [t] and [d] are realisations of the same phoneme /t/. The preattentive brain responses (mismatch negativity, MMN) to such stimuli from subjects from these two linguistic groups differ (Kazanina, Phillips, & Idsardi, 2006). In addition to VOT (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Kinney, & Lane, 1961), CP has been reported for various other features of speech, such as place of articulation (Johnson & Ralston, 1994; Liberman et al., 1957), manner of articulation (Fitch, Halwes, Erickson, & Liberman, 1980; Miller & Eimas, 1977), nasality of consonants (Larkey, Wald, & Strange, 1978; Miller & Eimas, 1977), vowel duration (Fujisaki, Nakamura, & Imoto, 1975; Minagawa-Kawai et al., 2007), and lexical tone (Francis, Ciocca, & Ng, 2003; Hallé, Chang, & Best, 2004; Peng et al., 2010; Wang, 1976; Xu, Gandour, & Francis, 2006).

Thousands of the world's languages are tone languages, using pitch contour, or its acoustic correlate, fundamental frequency (F_0), to distinguish lexical items (Yip, 2002). For example, in Mandarin, a variety of Chinese spoken throughout China, the syllable [ma] has the meaning "mother" when produced with a high level pitch contour but the meaning "hemp" when produced with a high rising pitch contour (Wang, 1973). Mandarin has an inventory of four lexical tones[1]: a high level tone (Tone 1); a high rising tone (Tone 2); a falling-rising (or dipping) tone (Tone 3); and a high falling tone (Tone 4). In contrast, Cantonese, which is spoken mostly in southern China, including Hong Kong, has an inventory of six lexical tones[2]: a high level tone (Tone 1); a high rising tone (Tone 2); a middle level tone (Tone 3); a low falling tone (Tone 4); a low rising tone (Tone 5); and a low level tone (Tone 6). Peng (2006) has studied the production of each tone in the Mandarin and Hong Kong Cantonese (henceforth, just Cantonese) tone systems in terms of both F_0 height and F_0 slope. Both the high level tone and high rising tone are acoustically similar in Mandarin and Cantonese, their speaker-normalised F_0 height and slope being almost identical. The tone system of Cantonese is acoustically denser than that of Mandarin. Peng (2006) found that the four Mandarin tones tend to be produced distinctly from each other, allowing the Mandarin listener to discriminate them readily. However, he observed significant overlap in the values of F_0 height and slope for the Cantonese tones. In particular, Tones 3 and 6 have the same F_0 slope and only slightly different F_0 height. Furthermore, Tones 2 and 5 have similar F_0 height and only slightly different F_0 slope. Although Wang (1971, p. 285) comments that "a fundamental principle is that the sounds of a language tend to maximise the phonetic distance from each other", Peng's (2006) observations appear to

[1] Mandarin also has a neutral tone, but this tone does not provide lexical contrast.

[2] Cantonese also has three entering tones on syllables with stop endings. The entering tones, referred to as Tones 7, 8, and 9, have similar pitch contours to Tones 1, 3, and 6, respectively, but have shorter duration.

provide a counterexample to this principle, and imply that the Cantonese listener might be required to make finer distinctions in perception of F_0 height and slope in order to discriminate certain tones than the Mandarin listener.

Cross-linguistic studies have shown that long-term language experience can influence how tones are perceived. For example, tone language speakers have greater CP to native tone contrasts than nontone language speakers: Mandarin vs. English (Bent, Bradlow, & Wright, 2006; Klein, Zatorre, Milner, & Zhao, 2001; Mattock & Burnham, 2006; Wang, 1976; Xu, Gandour, & Francis, 2006); Mandarin vs. French (Hallé et al., 2004); Thai vs. English (Gandour, Wong, & Hutchins, 1998). Moreover, native experience of one tone language can facilitate the perception of the lexical tones of other languages to some extent (Kaan, Barkley, Bao, & Wayland, 2008; Lee, Vakoch, & Wurm, 1996). However, few studies have compared explicitly the influence of language experience on the perception of tone contrasts that are native to two or more tone systems, for example, the high rising and high level tones that exist in both the Cantonese and Mandarin tone systems. Gandour (1983) has contrasted Cantonese and Mandarin listeners' responses to native tones, observing that Mandarin listeners make more errors than Cantonese listeners in distinguishing the high level tone and high rising tone, a result that he attributed to the different tone sandhi rules in these two languages.

Some studies have suggested that listeners with different language experiences pay more attention to different dimensions of the pitch contours. For example, Tse (1978) and Vance (1977) suggest that F_0 height is more important than F_0 slope for identification and acquisition of lexical tones in Cantonese. Using an INDSCAL multidimensional scaling analysis, Gandour (1983) found that Mandarin listeners place more emphasis on tone height than Thai listeners. This different sensitivity to different dimensions of pitch contours has also been exhibited by some electrophysiological data. A cross-language MMN study by Kaan et al. (2008) has provided evidence that native Thai listeners are more sensitive to late F_0 contour differences (mid-level vs. high rising) than Mandarin and English listeners, while English listeners are more sensitive to early F_0 contour differences (mid-level vs. low falling) than Thai and Mandarin listeners. Particularly interesting is their observation that sensitivity to early F_0 contour differences is suppressed by tone categorisation training. The above evidence all suggests that language experience influences the perception of lexical tones and, moreover, that different language experiences may drive listeners to pay more attention to different dimensions of the pitch contour. However, in the study by Kaan et al. (2008), the speech stimuli that were used were native—and therefore meaningful—to the Thai listeners, but not to the Mandarin or English listeners, a factor that influenced their perception.

In addition to linguistic experience, another factor influencing tone perception that has aroused much interest is the role of lexical information.

There is ample evidence that nonspeech is perceived categorically when it exhibits some of the critical features of speech (e.g., tone onset time: Holt, Lotto, & Diehl, 2004; noise with rapidly changing amplitude envelope: Mirman, Holt, & McClelland, 2004, see also the review by Repp, 1984). In terms of behavioural response, tone language speakers are better able to discriminate the pitch contours associated with distinct native tones in speech context than in nonspeech context (Lee et al., 1996; Xu, Gandour, Talavage et al., 2006). However, the speech context provides little advantage to either nontone-language speakers or non-native tone-language speakers to discriminate such pitch contours, and may even reverse the pattern (Lee et al., 1996; Xu, Gandour, Talavage et al., 2006).

This selective enhancement to tone perception in speech context is not evident in the early brain responses to tonal stimuli. Studies of the frequency following response indicate that neural plasticity for pitch representation at the level of the brainstem is dependent upon specific dimensions of pitch contours rather than speech per se (Krishnan & Gandour, 2009; Krishnan, Swaminathan, & Gandour, 2009). Studies have also shown that language experience can be reflected in a later event-related brain potential (ERP) component, the MMN (Näätänen, 2000, 2003). MMN is typically elicited by an oddball paradigm (Näätänen, Paavilainen, Rinne, & Alho, 2007) in which the subject's attention is not required. Cheour et al. (1998) and Näätänen et al. (1997) have demonstrated that non-native phonological distinctions elicit lower amplitude MMN than native phonological distinctions. CP of vowels and consonants is reflected in larger MMN amplitude for across-category deviants than within-category deviants (Kasai et al., 2001; Kazanina et al., 2006; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005). However, other researchers (e.g., Maiste, Wiens, Hunt, Scherg, & Picton, 1995; Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000) have shown that distinct stimuli from the same phonological category may nevertheless elicit an MMN. Although MMN, and perhaps earlier components too, can reflect linguistic experience, this experience-dependent plasticity is not specific to speech in the perception of pitch contours (Chandrasekaran, Krishnan, & Gandour, 2007, 2009).

In contrast, at later, attentive stages of processing, experience-driven effects are highly sensitive to categorical representations (Frenck-Mestre, Meunier, Espesser, Daffner, & Holcomb, 2005; Maiste et al., 1995). The P300 ERP component (Sutton, Braren, Zubin, & John, 1965) is also typically elicited by an oddball paradigm, similar to that used in MMN studies. However, modulation of the P300 component typically requires the subject to attend to the target stimuli. Some studies have suggested that P300 reflects the discrimination of speech stimuli based on phonological information (Frenck-Mestre et al., 2005; Maiste et al., 1995). In contrast, the MMN evoked by changes in speech sounds may indicate the detection of experience-relevant acoustic features (Chandrasekaran et al., 2007, 2009;

Maiste et al., 1995). In the oddball paradigm, regardless of attention, the brain must form a representation of the repetitive aspects of auditory stimuli before the occurrence of a deviant stimulus. Because the elicitation of the MMN and P300 components has different attentional requirements, they index different stages of brain processing. MMN reflects the automatic comparison of stimuli in the preattentive stage, no matter whether in the acoustic domain or the phonemic domain (Näätänen et al., 2007). In the active oddball paradigm, the participant builds up a mental schema of stimulus context as successive standards are presented (Donchin, 1981; Donchin & Coles, 1988; Polich, 2007). This memory schema must be updated whenever a new stimulus is detected. The resultant parietal P300 is modulated by the participant's overall arousal level as resources are allocated to process the stimulus—the less demanding the task, the greater the amplitude of the P300 (Polich, 2007).

P300 has been used to gauge the influence of linguistic experience on speech perception. For example, whereas spoken English contrasts the phonemes /r/ and /l/, Japanese does not. This difference in phonemic contrast is reflected in the limited or, for some individuals, totally absent ability of adult, native Japanese listeners to discriminate [r] and [l] sounds, both in terms of their behavioural responses and as indicated by their P300 responses (Buchwald, Guthrie, Schwafel, Erwin, & Vanlancker, 1994). Frenck-Mestre et al. (2005) have shown that the P300 response elicited by presentation of non-native deviant and native standard phonemes in an oddball paradigm depends on how listeners categorise the deviant within their native phonological system. In their experiments, the non-native (American English) deviant vowel /ɪ/ was perceived variously as /e/, /ø/, /æ/, or /ɛ/, but not as /ɪ/. Hence, for standard /ɛ/, the deviant /i/ was sometimes perceived as the standard, eliciting no significant difference in P300 amplitude. For standard /i/, however, the deviant was always perceived as a distinct phoneme from the standard, eliciting significantly increased P300 amplitude. Elicitation of the P300 component is not obligatory (Fabiani, Gratton, & Federmeier, 2007), hence these results suggest that modulation of P300 amplitude in response to linguistic stimuli provides a sufficient, but not necessary, index of phonological discrimination (Frenck-Mestre et al., 2005).

As observed above, the tone system of Cantonese is acoustically denser than that of Mandarin, requiring the Cantonese listener to make finer distinctions in perception of both F_0 height and slope in order to discriminate certain tones. Because lexical tones are defined primarily in terms of pitch contour, it is natural to ask whether native tone language listeners respond to the pitch contours of nonspeech sounds in the same way as to the pitch contours of tonal syllables. Brain-imaging data from tone perception studies have revealed that distinct brain regions are involved when pitch contours are superimposed on linguistic (pseudoword) or nonlinguistic (hum) carriers (Gandour et al., 2002). In that study, native Thai speakers showed left-lateralised inferior prefrontal activation when they discriminated linguistic tones compared to nonlinguistic pitches. The inferior prefrontal activation may reflect higher-level cognitive processes implicated in the extraction of phonological elements (Gandour et al., 2002). This finding suggests that the cognitive processing involved in pitch perception depends on whether or not the stimulus is perceived in speech context, irrespective of whether the stimulus bears meaning.

In the present experiment, we examine participants' behavioural and electrophysiological responses, obtained by electroencephalography (EEG), to speech and nonspeech (complex tone) stimuli. We present stimuli in an active oddball paradigm (e.g., Donchin, Ritter, & McCallum, 1978), randomly interspersing standards—having the pitch contour of either a high rising tone or a high level tone—with deviants of two types: a within-category deviant—an exemplar of the same linguistic category as the standard, but with different pitch contour—and an across-category deviant—an exemplar of a distinct linguistic category from the standard (for details, see *Stimuli* and *Procedure*, below). Native Mandarin and Cantonese participants' identification frequencies, reaction times, sensitivity indices (d' scores), and electrophysiological responses are analysed to determine which types of deviant—within-category or across-category—in which contexts—speech or nonspeech—they are able to process with greater ease. In particular, we consider the evidence in favour of the following hypothesis: that the amplitude of the P300 component elicited from Cantonese listeners in response to perception of deviant stimuli is greater than that elicited from Mandarin listeners, reflecting the greater ease with which Cantonese listeners engage phonological processing in order to distinguish the F_0 contours of distinct tone categories.

MATERIALS AND METHODS

Participants

Twenty-seven right-handed participants from The Chinese University of Hong Kong, with normal hearing and no reported history of neurological illness, were paid to take part in the experiment. The participants comprised 14 native Mandarin speakers (7 female; 7 male; mean age: 23.8 ± 3.8; Laterality Quotient: 75.9 ± 25.7 (Oldfield, 1971)) and 13 native Hong Kong Cantonese speakers (8 female and 5 male; mean age: 21.5 ± 2.2; Laterality Quotient: 71.0 ± 29.8). The average duration of residence of the Mandarin-speaking participants in Hong Kong was 1.38 ± 0.70 years; all had low proficiency in Cantonese. The Cantonese-speaking participants all had mid-level proficiency in Mandarin. No participant of either group had learned the first language of the other group before the age of seven. Participants were screened for musical experience; no participant had received musical training before the age of seven. Informed written consent was obtained from each participant. Approval to conduct the experiment was obtained from the Survey and Behavioural Research Ethics Committee of The Chinese University of Hong Kong.

Stimuli

The syllable [i] was produced with high level pitch contour by a male, native Mandarin speaker and recorded.[3] From the recorded utterance, 11 speech stimuli, each of duration 500 ms, were synthesised using Praat (Boersma & Weenink, 2009), with the pitch contours manipulated as illustrated in Figure 1. These pitch contours formed bilinear approximations (Wang, 1976) of the high level (Tone 1) and high rising (Tone 2) tones in both Mandarin and Cantonese. All 11 speech stimuli had the same formants: $F_1 = 280$ Hz (bandwidth 72 Hz); $F_2 = 2{,}500$ Hz (bandwidth 99 Hz); $F_3 = 3{,}120$ Hz (bandwidth 241 Hz); $F_4 = 4{,}088$ Hz (bandwidth 738 Hz) (all values averaged over the 150–350 ms segment). Eleven additional, nonspeech stimuli were synthesised from a complex tone (saw wave) with the same

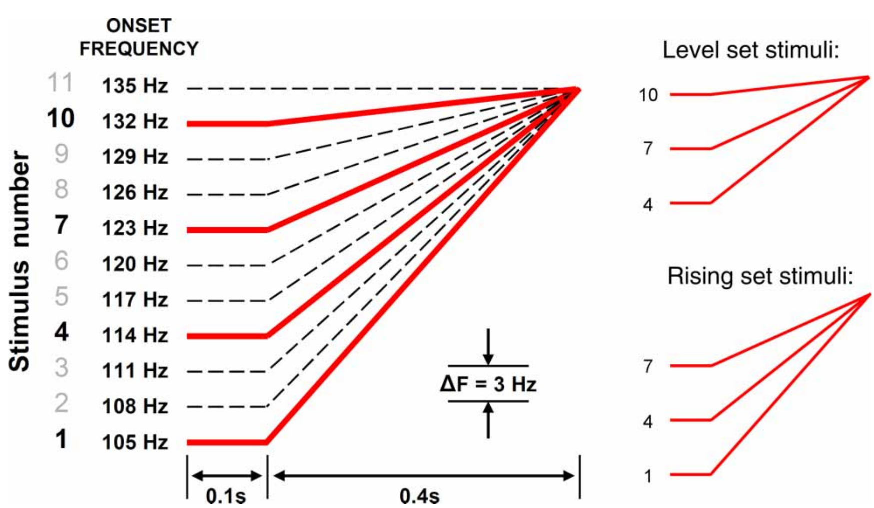

Figure 1. Pitch contours of the stimuli. [To view this figure in colour, please visit the online version of this Journal.]

[3] In a follow-up test, five Mandarin listeners and three Cantonese listeners were instructed to rate the naturalness of the speech stimuli on a 5-point scale, '1' indicating least natural, '5' indicating most natural. The Mandarin listeners rated the stimuli to have naturalness 3.89 ± 0.27 (mean $\pm$ standard error); the Cantonese listeners rated the stimuli to have similar naturalness, 3.90 ± 0.35.

pitch contours as the speech stimuli. Based on the subjective judgements of two independent researchers, the speech and nonspeech stimuli were presented at 65 dB and 75 dB, respectively, so that participants would perceive the two types of stimuli as having comparable loudness. The intensity envelopes of the two stimulus sets were closely matched and were kept constant across each continuum.

Procedure

Both the speech and nonspeech stimuli were organised into two stimulus sets: a *rising set* and a *level set*, shown in Figure 1. Each rising set (both speech and nonspeech) comprised stimuli #1, #4, and #7, and each level set comprised stimuli #4, #7, and #10. All stimuli were presented binaurally to participants via a pair of E·A·RTone 3A Insert Earphones (50Ω) manufactured by Etymotic Research Inc. The earphones were precalibrated by the manufacturer to provisional threshold sound pressure levels that are suitable for most purposes (Etymotic Research Inc., 2010; Wilber, Kruger, & Killion, 1988). No additional calibration was applied. For each rising set, stimulus #4 was presented as *standard*, stimulus #1 as *within-category deviant*, and stimulus #7 as *across-category deviant* in an oddball paradigm (e.g., Donchin et al., 1978). For each level set, stimulus #7 was presented as *standard*, stimulus #10 as *within-category deviant*, and stimulus #4 as *across-category deviant*, also in an oddball paradigm. Four hundred trials were presented for each stimulus set, comprising 75% standards, 12.5% within-category deviants, and 12.5% across-category deviants. Participants were instructed to press simultaneously both buttons of a mouse using their two thumbs whenever they heard a deviant of either type. The trials in each set were presented in pseudo-random order, with at least one standard preceding each deviant. The interval between successive trials was jittered in the range of 800–1,000 ms. Each stimulus set was presented in three blocks, with 2-minute breaks between blocks and 5-minute breaks between sets. The four stimulus sets (i.e., speech–rising, speech–level, nonspeech–rising, and nonspeech–level) were presented to participants in counterbalanced order. Throughout the experiment, EEG data were recorded as described in *Electrophysiological recordings*, below. A practice run for each stimulus set was presented before recording began in order to familiarise participants with the task and stimuli.

Participants then carried out a behavioural posttest in order to determine how they categorised the stimuli used in the main experiment. For both speech and nonspeech, stimuli #10 and #1 were presented to the participant before the task began as best examples of two categories that were labelled as "Sound 1" and "Sound 2", respectively—no reference was made to the tone categories of the speech stimuli. Participants were instructed to press the buttons on a keyboard to identify whether they perceived "Sound 1" or

"Sound 2". Speech and nonspeech stimuli were presented in separate blocks, comprising 15 presentations of each stimulus (#1, #4, #7, and #10) in pseudo-random order.

Electrophysiological recordings

EEG data were recorded using an Electrical Geodesics GES 250 system with 128-channel Ag/AgCl electrode arrays. Vertical eye movements were monitored by electrodes placed on the supra- and infraorbital ridges of each eye and horizontal eye movements by electrodes near the outer canthus of each eye. The recordings were referenced to the vertex electrode, filtered with an analogue 0.1–400 Hz band-pass filter, and digitised at a rate of 1,000 Hz. Electrode impedances were maintained below 50 kΩ, following the recommendation of the manufacturer (Electrical Geodesics, 2006a).

Data analysis

The hit rate (defined as the rate at which participants pressed the mouse button when a deviant of either type was presented), false alarm rate (defined as the rate at which participants pressed the mouse button when the standard was presented), and mean reaction time of each participant's behavioural responses were computed for each experimental condition of the main experiment. The sensitivity index, d', was calculated by subtracting the false alarm rate z-score from the hit rate z-score (Macmillan & Creelman, 1991). Three-way mixed design repeated-measures analysis of variance was carried out for both d' and reaction time with two within-subject factors, *category* (across-category vs. within-category) and *context* (speech vs. nonspeech), and one between-subject factor, *language* (Mandarin vs. Cantonese). The posttest identification frequencies for stimuli #1, #4, #7, and #10 were also calculated. Three-way mixed design repeated-measures analysis of variance was carried out on identification frequency with two within-subject factors, *stimulus* (stimulus #4 vs. stimulus #7) and *context* (speech vs. nonspeech), and one between-subject factor, *language* (Mandarin vs. Cantonese). All post hoc analyses were conducted according to the general guidelines for analysing effects in a three-factor design recommended by Maxwell and Delaney (2004, p. 377). The p-values of the t-tests that are reported were all corrected for multiple comparisons wherever appropriate. Unless stated otherwise, all tests of significance were conducted at $p < .05$.

The EEG recordings were re-referenced offline against average-mastoid reference, and low-pass filtered at 30 Hz. For each participant in each experimental condition, event-related potentials were calculated over a 900 ms interval time-locked to stimulus onset relative to a 100 ms prestimulus baseline. Trials with ocular artifacts were excluded from averaging to ERP, as were trials in which the participant incorrectly responded to a standard by making a button press. Statistical analyses were conducted on the amplitude and peak latency of the N1, P2, N2, and P300 components. Each component was determined from the maximum (or minimum) amplitude within a specific time window based on previous studies and confirmed by the grand-averaged ERP waves across all experimental conditions (see Figure 2). Analysis of each component was carried out on the difference wave between deviants of each type and standard. Electrode locations were constrained to the region of interest of each component based on previous studies and confirmed by the topographic distribution maps (see Figure 2). In particular, N1 was measured at eight anterior electrodes (numbered according to the 128-channel GSN 200 sensor layout; Electrical Geodesics, 2007) where N1 amplitude was expected to peak (Luck, 2005; Näätänen & Picton, 1987)—4, 11, 16, and 20 (10–10 system equivalents: F2, Fz, AFz, and F1, respectively), and 5, 10, 12, and 19 (no 10–10 system equivalents)—using a time window of width 60 ms centred on the peak latency within the range 50–150 ms. P2 was measured at six central electrodes where P2 amplitude was expected to peak (Alain & Tremblay, 2007; Luck, 2005)—55 and 129 (10–10 system equivalents: CPz and Cz, respectively), and 7, 32, 81, and 107 (no 10–10

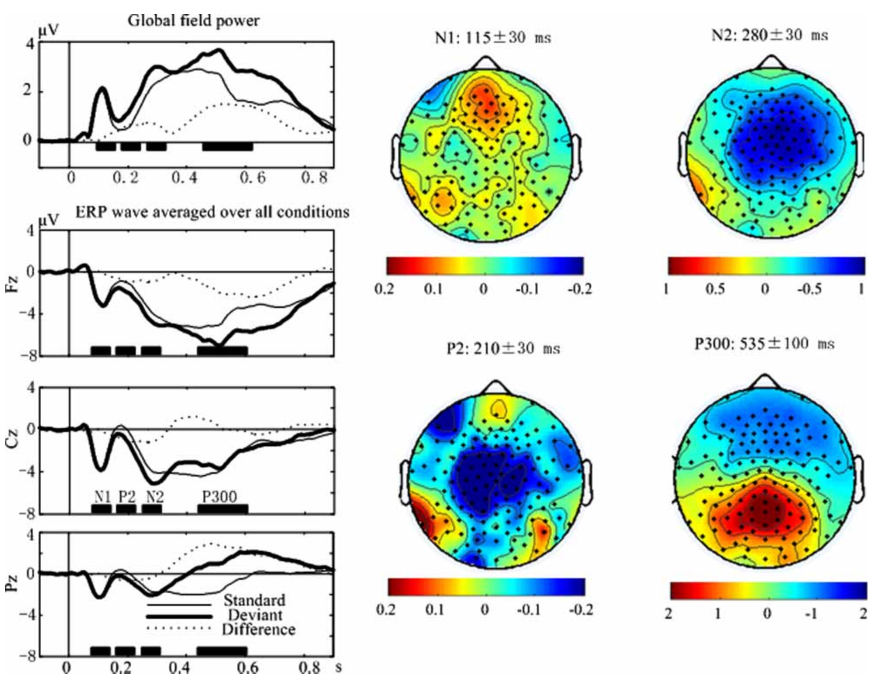

Figure 2. Global field power, ERP wave averaged over all experimental conditions, and topographic distribution maps for the N1, P2, N2, and P300 components. [To view this figure in colour, please visit the online version of this Journal.]

system equivalents)—using a time window of width 60 ms centred on the peak latency within the range 150–250 ms. N2 was measured at 11 fronto-central electrodes where N2 amplitude was expected to peak (Näätänen, 1982; Näätänen et al., 2007)—6, 55, and 129 (10–10 system equivalents: FCz, CPz, and Cz, respectively), and 5, 7, 12, 13, 32, 81, 107, and 113 (no 10–10 system equivalents)—using a time window of width 60 ms centred on the peak latency within the range 250–350 ms. P300 was measured at seven parieto-occipital electrodes where P300 amplitude was expected to peak (Polich, 2007)—61, 62, 68, and 79 (10–10 system equivalents: P1, Pz, POz, and P2, respectively), and 67, 73, and 78 (no 10–10 system equivalents)—using a time window of width 200 ms centred on the peak latency within the range 300–700 ms. (The electrode positions are summarised in the Supplementary Information, Figure S1.)

The amplitude of each ERP component was determined by calculating the adaptive mean amplitude (Electrical Geodesics, 2006b; Scerif, Worden, Davidson, Seiger, & Casey, 2006) over electrodes within the specified region of interest and time window. Difference waves were obtained by subtracting the standard ERPs from each of the deviant ERPs. Three-way mixed design repeated measures analysis of variance was conducted on both the adaptive mean amplitude of the difference wave obtained for each type of deviant and peak latency of the difference wave with two within-subject factors, *category* (across-category vs. within-category) and *context* (speech vs. nonspeech), and one between-subject factor, *language* (Mandarin vs. Cantonese) for both the rising set and level set stimuli.

RESULTS

Behavioural data

Identification frequency

The mean posttest identification frequencies for both speech and nonspeech stimuli are shown in Figure 3 for Mandarin and Cantonese participants. Participants consistently identified stimulus #4 as belonging to the same category as stimulus #1, i.e., "Sound 2", both for speech (Mandarin: 89.4%; Cantonese: 80.2%) and for nonspeech (Mandarin: 93.9%; Cantonese: 87.9%). Participants identified stimulus #7 as belonging to a distinct category, i.e., "Sound 1", both for speech (Mandarin: 89.9%; Cantonese: 97.5%) and for nonspeech (Mandarin: 67.8%; Cantonese: 73.0%). The analysis of variance indicated no significant language-related interaction effects, $p > .05$, implying that there was no significant difference in identification performance according to participants' language backgrounds. However, there was a significant stimulus $\times$ context interaction,

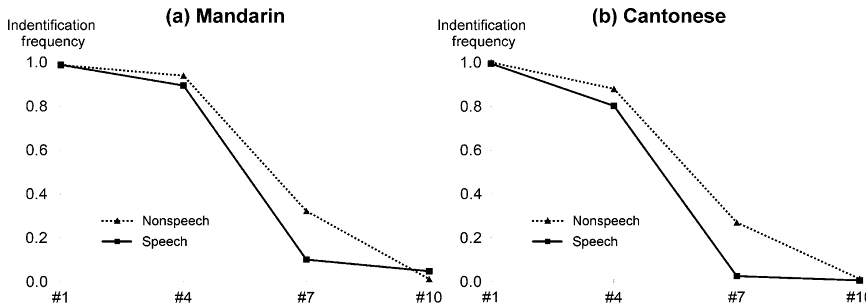

Figure 3. Identification frequencies for "Sound 2" for: (a) Mandarin participants and (b) Cantonese participants.

$F(1, 25) = 10.282$, $MSE = 0.229$, $p < .005$, as well as main effects of stimulus, $F(1, 25) = 291.970$, $MSE = 12.715$, $p < .001$, and context, $F(1, 25) = 29.574$, $MSE = 0.635$, $p < .001$.

According to the post hoc tests, stimulus #7 was identified more frequently as "Sound 1" in the speech context than in the nonspeech context, $t(26) = 4.193$, $p < .001$, but the identification frequency of stimulus #4 was found not to vary significantly in the speech and nonspeech contexts, $t(26) = 1.848$. In addition, stimulus #4 was identified more frequently than stimulus #7 as "Sound 2" for both speech stimuli, $t(26) = 16.741$, $p < .001$, and nonspeech stimuli, $t(26) = 11.813$, $p < .001$.

Sensitivity index: d′ score

The *d′* scores for the four stimulus sets elicited from both Mandarin and Cantonese participants are summarised in Figure 4. For the rising set (Figure 4a), the analysis of variance revealed significant main effects of context, $F(1, 25) = 6.439$, $MSE = 4.406$, $p < .05$, and category, $F(1, 25) = 19.052$, $MSE = 51.213$, $p < .001$. A significant context × category interaction was also observed, $F(1, 25) = 15.073$, $MSE = 4.799$, $p < .001$. No other effects were significant. In particular, there was no main effect or interaction involving language. In the post hoc tests, the nonspeech context was found to elicit greater *d′* scores than the speech context for within-category stimuli, $t(26) = 4.193$, $p < .001$, but not across-category stimuli, $t(26) = -0.0653$. In addition, across-category stimuli were found to elicit greater *d′* scores than within-category stimuli for both speech stimuli, $t(26) = 4.679$, $p < .001$, and nonspeech stimuli, $t(26) = 3.417$, $p < .005$.

For the level set (Figure 4b), the analysis of variance indicated significant main effects of context, $F(1, 25) = 5.3745$, $MSE = 3.281$, $p < .05$, and category, $F(1, 25) = 39.416$, $MSE = 62.427$, $p < .001$. There was also a significant language × context × category interaction, $F(1, 25) = 8.982$, $MSE = 4.976$,

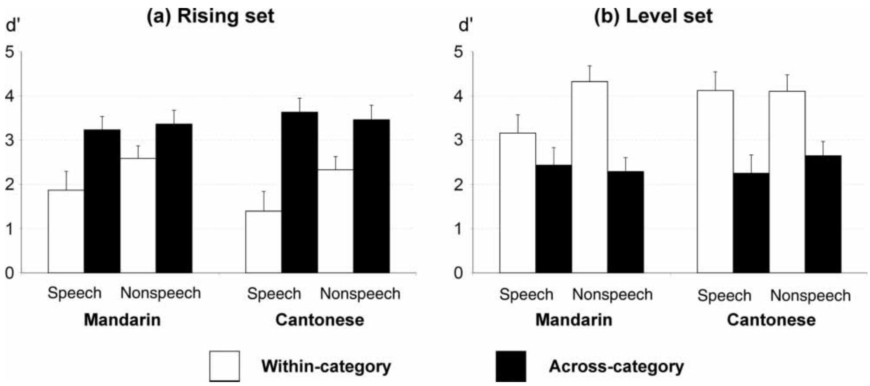

Figure 4. Sensitivity index (d' scores) for: (a) the rising set and (b) the level set.

$p < .01$. This was followed up by two-way analysis of d' for factors *category* (across-category vs. within-category) and *context* (speech vs. nonspeech) for both Mandarin and Cantonese participants.

For Mandarin participants, in addition to main effects of both context, $F(1, 13) = 6.853$, $MSE = 3.626$, $p < .05$, and category, $F(1, 13) = 14.181$, $MSE = 26.676$, $p < .01$, a context × category interaction, $F(1, 13) = 8.576$, $MSE = 5.991$, $p < .05$, was observed. Further analysis indicated that the nonspeech context elicited greater d' scores than the speech context for within-category stimuli, $t(13) = 4.479$, $p < .005$, but not across-category stimuli, $t(13) = -0.442$. Also, the d' scores for within-category stimuli exceeded those for across-category stimuli in the nonspeech context, $t(13) = 5.746$, $p < .001$, but not the speech context, $t(13) = 1.473$.

For Cantonese participants, however, there was no significant context × category interaction, $F(1, 12) = 1.375$, $MSE = 0.547$. A significant main effect of category, $F(1, 12) = 28.488$, $MSE = 35.946$, $p < .001$, was observed, but not context, $F(1, 12) = 0.662$, $MSE = 0.463$. Comparison of the marginal means indicated that the d' scores for within-category stimuli exceeded those for across-category stimuli.

Reaction time

Reaction times for trials in which participants made a behavioural response are presented in Figure 5 for both Mandarin and Cantonese participants. For the rising set, the analysis of variance revealed no significant effects (Figure 5a). For the level set (Figure 5b), however, there was a significant main effect of category $F(1, 25) = 2.134$, $MSE = 6227.541$, $p < .001$. No other effects were significant. In particular, there were no significant category × language or context × category × language interactions for either set. Thus, there was no evidence that either group of

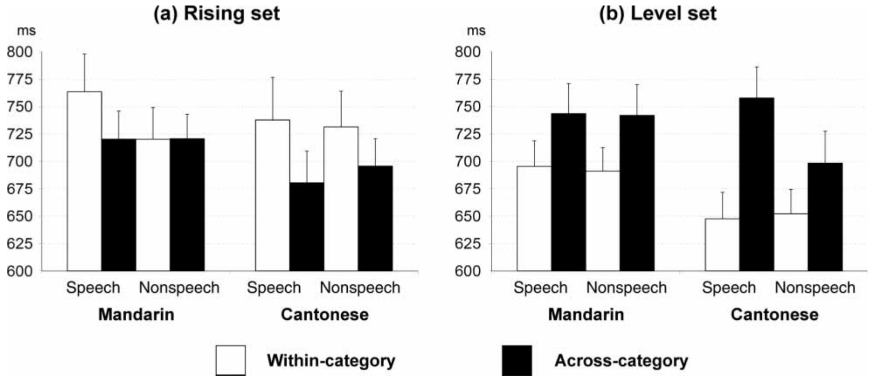

Figure 5. Reaction times for: (a) the rising set and (b) the level set.

participants, Mandarin or Cantonese, made a button press in response to deviant stimuli significantly more quickly than the other.

Electrophysiological data

The ERPs and corresponding difference waves elicited from both Mandarin and Cantonese participants are presented in Figure 6, for the rising set, and Figure 7, for the level set.

Analysis of the amplitude and peak latency of N1, P2, and N2 (see Supplementary Information for details) indicated that none of these early components was modulated by participants' language backgrounds. Likewise, language background was found not to modulate P300 peak latency (see Supplementary Information).

The analysis of variance of P300 amplitude for the rising set (Figure 8a) revealed a significant context × category × language interaction, $F(1, 25) = 5.474$, $MSE = 13.628$, $p < .05$. Significant context × category, $F(1, 25) = 4.630$, $MSE = 11.527$, $p < .05$, and category × language interactions, $F(1, 25) = 7.983$, $MSE = 35.402$, $p < .01$, as well as a significant main effect of category, $F(1, 25) = 18.098$, $MSE = 80.259$, $p < .005$, were also observed. The significant three-way interaction indicates the presence of a CP effect, with the electrophysiological response modulated by both context and participants' language backgrounds. The three-way analysis of P300 amplitude was followed up by two-way analysis of factors *category* (across-category vs. within-category) and *context* (speech vs. nonspeech) for both Mandarin and Cantonese participants.

For Mandarin participants, no main effects or interactions were found to be significant. For Cantonese participants, there were no significant main effects, but a context × category interaction, $F(1, 12) = 6.983$, $MSE = 24.215$,

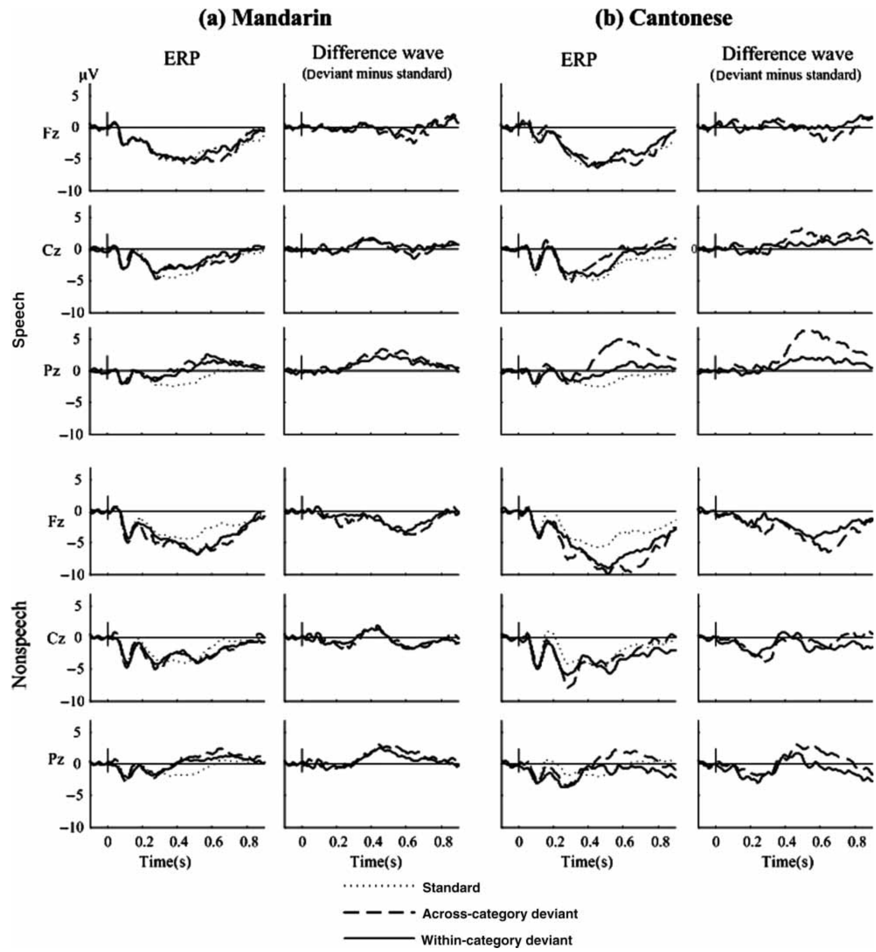

Figure 6. Electrophysiological data for the rising set elicited from: (a) Mandarin participants and (b) Cantonese participants. Each panel shows: (left column) the ERPs elicited by across-category deviants, within-category deviants, and standards; and (right column) the difference waves for across-category deviants minus standard and within-category deviants minus standard. Waveforms are displayed for three representative electrodes: 11, 129, and 62 (10–10 system equivalents: Fz, Cz, and Pz, respectively).

$p < .05$, was observed. Further analysis indicated that across-category stimuli elicited significantly greater P300 amplitude than within-category stimuli in both the speech context, $t(12) = 4.066$, $p < .01$, and the nonspeech context, $t(12) = 2.604$, $p < .05$. There was no significant difference in P300 amplitude between speech and nonspeech contexts for either within-category or across-category stimuli.

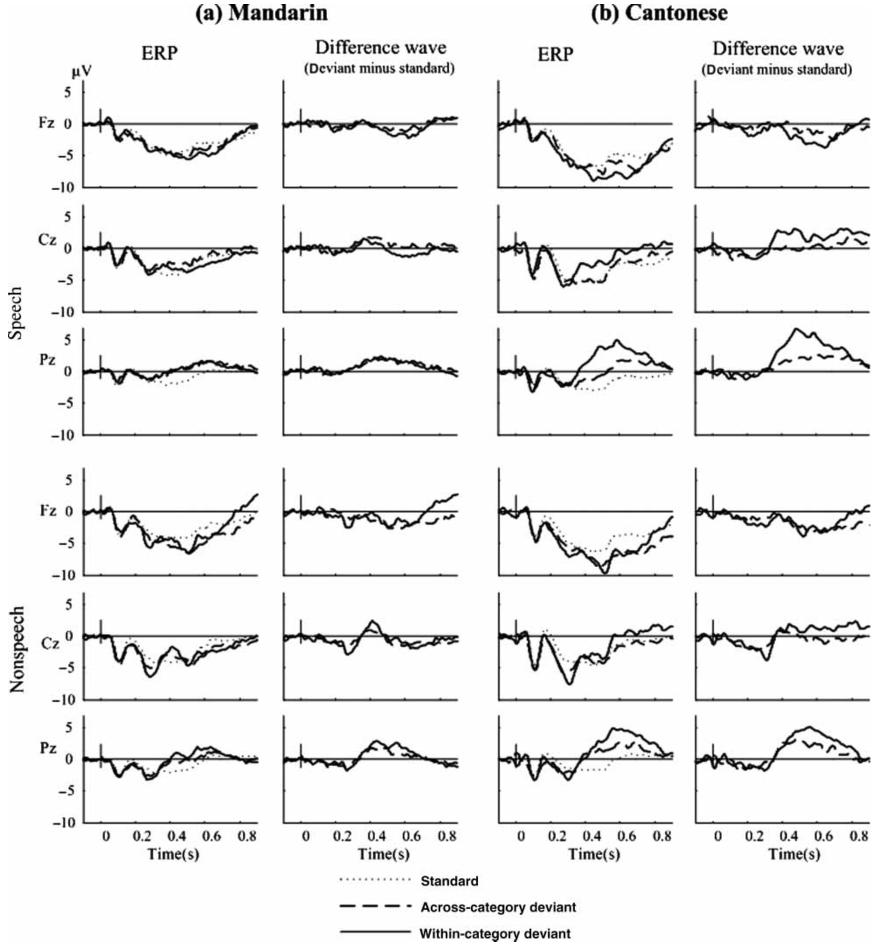

Figure 7. Electrophysiological data for the level set elicited from: (a) Mandarin participants and (b) Cantonese participants. Each panel shows: (left column) the ERPs elicited by across-category deviants, within-category deviants, and standards; and (right column) the difference waves for across-category deviants minus standard and within-category deviants minus standard. Waveforms are displayed for three representative electrodes: 11, 129, and 62 (10–10 system equivalents: Fz, Cz, and Pz, respectively).

For the level set, the analysis of variance of P300 amplitude (Figure 8b) indicated significant main effects of language, $F(1, 25) = 8.122$, $MSE = 90.279$, $p < .01$, and category, $F(1, 25) = 8.122$, $MSE = 90.279$, $p < .01$, but not context, $F(1, 25) = 0.229$, $MSE = 1.974$, as well as a significant category $\times$ language interaction, $F(1, 25) = 12.641$, $MSE = 48.196$, $p < .005$. Analysis of the marginal means indicated that within-category stimuli

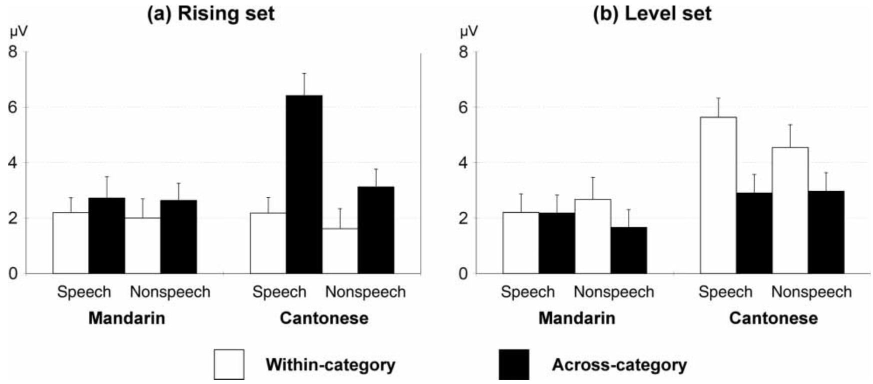

Figure 8. P300 amplitude elicited by: (a) the rising set and (b) the level set. P300 amplitude was calculated from the average voltage measured at electrodes 61, 62, 68, and 79 (10–10 system equivalents: P1, Pz, POz, and P2, respectively) and 63, 73, and 78 using a time window of width 200 ms centred on peak latency within the 300–700 ms range at each electrode.

elicited greater P300 amplitude than across-category stimuli for Cantonese participants, $t(12) = 3.691$, $p < .005$, but not for Mandarin participants, $t(13) = 1.079$. Additional independent samples t-tests showed that greater P300 amplitude was elicited from Cantonese participants than from Mandarin participants for within-category stimuli, $t(25) = 3.418$, $p < .005$, but not across-category stimuli, $t(25) = 1.421$.

DISCUSSION

Language background did not modulate either participants' abilities to correctly identify the category of stimuli in the posttest or their reaction times in the main experiment. For the rising set, language background was found not to modulate participants' d' scores, which provide simple measures of their abilities to discriminate deviant stimuli from standard. Participants, regardless of language background, responded categorically to both the speech and nonspeech stimuli. There was no significant difference in across-category discriminability for speech and nonspeech. However, within-category discriminability was greater for nonspeech than for speech due to the speech stimuli representing the same phonological category. These results are all consistent with participants having stronger CP of speech than nonspeech in their native language, as found previously (e.g., Lee et al., 1996; Xu, Gandour, & Francis, 2006).

For the level set, regardless of context or language, participants' d' scores showed that within-category discriminability was greater than across-category discriminability. In other words, stimulus #7 was categorised into the same category as stimulus #4, both stimuli having rising pitch contour. This shift on the category boundary,[4] from between stimuli #4 and #7 for the rising set to between stimuli #7 and #10 for the level set, may be due to the global sequential (range–frequency) effect (Repp & Liberman, 1987).[5] In addition to the boundary shift, language background was found to modulate participants' *d′* scores, there being a significant language × context × category interaction. For Mandarin participants, within-category discriminability was greater for nonspeech than for speech. There was no significant difference in discriminability for the two speech deviants, hence the *d′* data provide no evidence of CP of level set speech stimuli by Mandarin participants.

However, the global sequential (range–frequency) effect does not provide a full explanation for the observed data: in the nonspeech condition, within-category stimulus #10 was more easily discriminated from standard than across-category stimulus #4. We propose that psychophysical processing, for both speech and nonspeech stimuli, was engaged, enhancing the perception of the within-category deviant stimulus #10 as a level pitch contour, thereby allowing Mandarin participants to discriminate it more easily from the standard stimulus #7, having rising pitch contour. We argue that the psychophysical processing was engaged more strongly for nonspeech stimuli than for speech stimuli. Infant studies (e.g., Mattock & Burnham, 2006) have shown that perceptual reorganisation through experience is more salient for speech stimuli than for nonspeech stimuli. The discriminability of non-native phonological contrasts declines significantly with age for speech stimuli, but less so for nonspeech stimuli. Evidence for such a psychophysical boundary between level and rising pitch contours has been reported previously for both nontone language speakers and phonetically trained tone language speakers (Wang, 1976).

For Cantonese participants, a slightly different pattern was observed in the *d′* scores. There was no category × context interaction effect, only a main category effect, with higher discriminability between stimuli #7 and #10 than between stimuli #7 and #4. In addition to the global sequential

[4] Two groups (10 subjects per group) of Mandarin speakers participated in a follow-up experiment. One group was instructed to identify stimuli within the rising set (stimuli #1–#7) while the second group was instructed to identify stimuli within the level set (stimuli #4–#10). The categorical boundary for the rising set was observed to lie between #4 and #5 for both speech and nonspeech, while the boundary for the level set lay between #7 and #8 for speech, confirming the presence of a range–frequency effect—for nonspeech, the position of the boundary was slightly higher at stimulus #8.

[5] We will continue to use the terms across-category and within-category to refer to the category memberships obtained from participants during the forced-choice identification test described in Procedure irrespective of changes in category boundary position due to the range–frequency or other effects.

(range–frequency) effect (Repp & Liberman, 1987), the observed shift in category boundary may again be due, in part, to the presence of a psychophysical boundary near stimulus #10. However, another possibility is that some Cantonese participants perceived stimuli #10 and #7, not as Tone 1 or Tone 2, but as Tone 3 (middle level tone) and Tone 5 (low rising tone), respectively. In a previous study (Peng et al., 2010), in which both Mandarin and Cantonese listeners were required to discriminate the same speech stimuli on the high rising (stimulus #1) to high level (stimulus #11, see Figure 1) continuum, Mandarin and Cantonese listeners showed different boundary patterns. For Mandarin listeners, there was a single peak in the two-step discrimination curve located at pair #6 and #8. In contrast, for Cantonese listeners, there were two peaks, one located at pair #6 and #8, the other at pair #9 and #11. This additional peak at the level end of the discrimination curve for Cantonese listeners might reflect a secondary phonological boundary, a proposal that will be further investigated in the future.

We now consider the electrophysiological data. The amplitudes and peak latencies of the early ERP components—N1, P2, and N2—were modulated by both stimulus category and context, but not by the language backgrounds of participants (see Supplementary Information). However, P300 amplitude was modulated by language background. In the oddball paradigm, a participant builds up a schema of the representative features of the standard before occurrence of a deviant. In a passive oddball paradigm, the representative features can be built up even though the participant does not attend to the stimuli, in which case an MMN is elicited. These features may relate to memory traces shaped by linguistic experience, but not necessarily to phonemic categories (Rivera-Gaxiola et al., 2000). Moreover, these features may not be constrained to speech context (Chandrasekaran et al., 2009). In contrast, in the active oddball paradigm, the participant pays focal attention to the stimuli. According to the context-updating hypothesis (Donchin, 1981; Donchin & Coles, 1988), P300 amplitude reflects the ease with which the participant updates mental schemata of stimulus context in response to changes in stimulus attributes (Polich, 2007), particularly phonemic attributes (Maiste et al., 1995). Frenck-Mestre et al. (2005) have shown that the P300 component elicited by speech stimuli indexes phonological discrimination, with P300 amplitude being greater for deviants that are perceived as phonologically distinct from the standard. Deviants that differ only phonetically elicit no significant change in P300 amplitude. Therefore, for within-category deviants, we expect little or no modulation of P300 amplitude by context, since neither speech nor nonspeech within-category deviants provide a phonological contrast by which memory schemata of stimulus context need be updated. However, for across-category stimuli, we expect significantly greater P300 amplitude to be elicited from speech stimuli, which are phonologically contrastive with the standard, than from nonspeech stimuli, which are not phonologically contrastive. In other words, discriminations that are made based on phonological processing should be expected to modulate P300 amplitude, but discriminations that are made based on psychophysical processing should not.

For the rising set, different P300 amplitudes were elicited from the two groups of participants. For Mandarin participants, none of the experimental conditions was found to significantly modulate P300 amplitude, although there was a nonsignificant trend for across-category deviants to elicit higher amplitude than within-category deviants. This contrasts with the result of the d' analysis, in which Mandarin participants were found to discriminate across-category stimuli from standard more easily than within-category stimuli. For Cantonese participants, however, across-category deviants did elicit significantly higher amplitude than within-category deviants, particularly in the speech context. This is consistent with the result of the d' analysis, in which Cantonese participants were found to discriminate across-category stimuli from standard more easily than within-category stimuli. Thus, we observe a mismatch in the electrophysiological and behavioural responses for Mandarin participants, but not Cantonese participants. This mismatch may be because the Mandarin participants, unlike the Cantonese participants, failed to make significant use of the information contained in the pitch contour to engage phonological processing as indexed by P300 amplitude. Both groups of participants were nevertheless able to identify behaviourally the categories of tonal stimuli equally accurately and rapidly. Behavioural measures, such as d' scores, reflect the results of *all* the cognitive processes involved in stimulus discrimination.

For the level set, the P300 amplitudes elicited from the two groups of participants differed significantly. These data indicate that, just as for the rising set, Cantonese participants engaged phonological processing more strongly than Mandarin participants. However, the deviant stimulus #4 elicited smaller P300 amplitude than deviant stimulus #10, a result that we attribute, in part, to the shift of category boundary due to the range–frequency effect (Repp & Liberman, 1987). For Mandarin participants, based on the behavioural data, we have argued that stimuli were discriminated primarily as a result of psychophysical processing. The P300 data for Mandarin participants are consistent with this: engagement of phonological processing in order to discriminate stimuli is too weak to elicit greater P300 amplitude for one type of deviant than the other. For Cantonese participants, the electrophysiological and behavioural data are also consistent. However, because of the presence of a secondary phonological category boundary between stimuli #7 and #10, phonological processing is engaged in order to discriminate these stimuli, thereby increasing P300 amplitude (Frenck-Mestre et al., 2005).

In the present experiment, the level set and rising set used different standards. Although these two standards were identified as high rising tone (#4 in rising set) and high level tone (#7 in level set), respectively, with high frequency, the standard that was used in the level set is not a prototypical level tone. As a result, the identification frequency of the level tone standard was lower than that of the rising tone standard. In an oddball paradigm, if nonprototypical stimuli are used as standard and prototypical stimuli used as deviant, participants are less able to establish a representative phonological category in the mental schema (Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987; Rivera-Gaxiola et al., 2000). Consequently, the ERP elicited from the deviant will more strongly reflect psychophysical processing. In the experiment reported here, this engagement of psychophysical processing in the level set is stronger for Mandarin speakers.

The results of the present experiment are consistent with our hypothesis that Cantonese listeners distinguish phonologically contrastive differences in pitch contour with greater ease than Mandarin listeners. This hypothesis may be tested directly in future research by comparing the difference limens for different pitch contours elicited from Mandarin and Cantonese listeners (cf. Hall & Plack, 2009).

Manuscript received 26 March 2010
Revised manuscript received 23 August 2010
First published online day 23 May 2011

REFERENCES

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*(3), 197–207.

Alain, C., & Tremblay, K. (2007). The role of event-related brain potentials in assessing central auditory processing. *Journal of the American Academy of Audiology*, *18*, 573–589.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103.

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved from http://www.praat.org/

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., & Vanlancker, D. (1994). Influence of language structure on brain-behavior development. *Brain and Language*, *46*(4), 607–619.

Burnham, D. K., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. *Journal of Child Language*, *18*(2), 231–260.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport*, *18*(18), 1963–1967.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, *16*(8), 1375–1387.

Donchin, E. (1981). Surprise! … surprise? *Psychophysiology*, *18*(5), 493–513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In P. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain-event related potentials in man* (pp. 349–411). New York: Academic Press.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Electrical Geodesics. (2006a). GES Hardware Technical Manual, S-MAN-200-COMR-001, December 21, 2006. Eugene, OR: Author.

Electrical Geodesics. (2006b). Net Station Viewer and Waveform Tools Tutorial, S-MAN-200-TVWR-001, December 21, 2006. Eugene, OR: Author.

Electrical Geodesics. (2007). Geodesic Sensor Net Technical Manual, S-MAN-200-GSNR-001, January 31, 2007. Eugene, OR: Author.

Etymotic Research Inc. (2010). ER·3A Data Sheet. Retrieved from http://www.etymotic.com/pdf/er3a-datasheet.pdf

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology 3rd edition* (pp. 85–119). Cambridge, New York: Cambridge University Press.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. (1980). Perceptual equivalent of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, *27*(4), 343–350.

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

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(9), 3221–3225.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1496–1510.

Fujisaki, H., Nakamura, K., & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 197–219). London: Academic Press.

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

Gandour, J. T., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport*, *9*(9), 2115–2119.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y. X., & Li, X. J. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience*, *14*(7), 1076–1087.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, *61*(1), 191–218.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, *19*(3), 576–585.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, *116*(3), 1763–1773.

Johnson, K., & Ralston, J. V. (1994). Automaticity in speech perception: Some speech/nonspeech comparisons. *Phonetica*, *51*(4), 195–209.

Kaan, E., Barkley, C., 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*(1), 53–69.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., et al. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *Neuroreport*, *12*(11), 2467–2471.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11381–11386.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, *13*(4), 646–653.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, *110*(3), 135–148.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, *21*(6), 1092–1105.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

Larkey, L., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, *23*(4), 299–312.

Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, *4*, 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*, 379–388.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

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 and Hearing*, *16*(1), 68–90.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *The Journal of the Acoustical Society of America*, *61*(3), 835–845.

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.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America*, *116*(2), 1198–1207.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection. *Psychological Bulletin*, *92*(3), 605–640.

Näätänen, R. (2000). Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology*, *37*(1), 3–10.

Näätänen, R. (2003). Mismatch negativity: Clinical research and possible applications. *International Journal of Psychophysiology*, *48*(2), 179–188.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590.

Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, *92*(1), 26–32.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 134–154.

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. *Journal of Phonetics*. Advance online publication. doi:10.1016/j.wocn.2010.09.003

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice*, *10*, 243–335.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge, UK: Cambridge University Press.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, *111*(1–2), 13–23.

Scerif, G., Worden, M. S., Davidson, M., Seiger, L., & Casey, B. J. (2006). Context modulates early stimulus processing when resolving stimulus-response conflict. *Journal of Cognitive Neuroscience*, *18*(5), 781–792.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188.

Tse, J. K.-P. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language*, *5*(2), 191–204.

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, *34*(2), 93–107.

Wang, W. S-Y. (1971). The basis of speech. In C. E. Reed (Ed.), *The learning of language* (pp. 267–306). New York: Appleton-Century-Crofts.

Wang, W. S-Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–60.

Wang, W. S-Y. (1976). Language change. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech* (Vol. 280, pp. 61–72). New York: New York Academy of Sciences.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63.

Wilber, L. A., Kruger, B., & Killion, M. C. (1988). Reference thresholds for the ER-3A insert earphone. *The Journal of the Acoustical Society of America*, *83*(2), 669–676.

Xu, Y. S., 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), 1063–1074.

Xu, Y. S., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y. X., et al. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173–183.

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.