J. Child Lang., Page 1 of 22. © Cambridge University Press 2016
doi:10.1017/S0305000916000581

The development of categorical perception of Mandarin tones in four- to seven-year-old children*

FEI CHEN

Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

GANG PENG

Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, and Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR

NAN YAN

Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

AND

LAN WANG

Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

(*Received* 16 *December* 2015 – *Revised* 29 *July* 2016 – *Accepted* 21 *October* 2016)

[*] This work was partially supported by grants from National Natural Science Foundation of China (NSFC: 61135003, 11474300, 91420301), Shenzhen Speech Rehabilitation Technology Laboratory, and Shenzhen Fundamental Research Program JCYJ20130401170306806. We thank all the child participants, from Guangming Central Kindergarten, Taoyuan Secondary Kindergarten, and Taoyuan Zhuoya Primary School, and their parents, for their participation and cooperation. Addresses for correspondence: Gang Peng, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China. tel: (+852) 3400 8462; fax: (+852) 2334 0185; e-mail: gpengjack@gmail.com. Nan Yan, Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. tel: (+86) 86392174; fax: (+86) 86392194; e-mail: nan.yan@siat.ac.cn

ABSTRACT

To track the course of development in children's fine-grained perception of Mandarin tones, the present study explored how categorical perception (CP) of Mandarin tones emerges along age among 70 four- to seven-year-old children and 16 adults. Prominent discrimination peaks were found for both the child and the adult groups, and they were well aligned with the corresponding identification crossovers. Moreover, six-year-olds showed a much narrower width (i.e. a sharper slope) compared to younger children, and have already acquired adult-like identification competence of Mandarin high-level and mid-rising tones. Although the ability to discriminate within-category tone pairs did not change, the between-category discrimination accuracies were positively correlated with chronological ages among child participants. We assume that the perceptual refinement of Mandarin tones in young children may be driven by an accumulation of perceptual development from the tonal information of the ambient sound input.

INTRODUCTION

Detailed research into the biological foundations of language in the context of growth and maturation – including babbling around six months of age to the production of full meaningful sentences by the age of three years (e.g. Kuhl, 2009; Saffran, Werker & Werner, 2006) – has led some to suggest that young children have innate dispositions and incredible abilities to learn any natural language (Lenneberg, 1967). Although newborns begin life with the ability to discriminate both native and non-native phonological contrasts attested in the world's languages (e.g. Best & Tyler, 2007; Eimas, Siqueland, Jusczyk & Vigorito, 1971; Polka & Werker, 1994; Streeter, 1976; Werker & Tees, 1984), their ability to discriminate non-native consonants and vowels gradually declines between 6 and 12 months as a result of consistent exposure to a specific language (e.g. Best & McRoberts, 2003; Best, McRoberts, LaFleur & Silver-Isenstadt, 1995; Werker & Tees, 1984), whereas similar ability to discriminate native consonants and vowels are maintained and enhanced over the same time period (Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006). The language-dependent perceptual reorganization during infancy with perceptual narrowing for the native phonological contrasts is considered to be largely attributable to 'statistical learning': tracking the frequencies of the exposed sound tokens and using this information to tune into native phonological categories (Best & McRoberts, 2003; Kuhl, 2004; Maye, Werker & Gerken, 2002).

Such phonological contrasts of different spoken languages can be divided into segmental units, including consonants and vowels, and suprasegmental

units, such as stresses and tones. Almost 60–70% of the world's languages are tone languages (Yip, 2002), and over half of the world's people speak a tone language (Fromkin, 1978). Many studies on infant tone perception have revealed that the process of perceptual reorganization is shaped by the ambient tonal language. In one longitudinal case study, the perceptual discrimination of Cantonese tones was found to begin at the tenth month (Tse, 1978). In another study by Mattock and Burnham (2006), infants were tested on their ability to discriminate Thai tone contrasts. Results indicated that English-exposed infants demonstrated attenuation in tone sensitivity between 6 and 9 months, while Chinese-learning infants exhibited a sustained sensitivity to Thai tones over this period. Moreover, developmental changes in tone perception were systematically explored by Yeung, Chen, and Werker (2013). Their results demonstrated that language experience might affect the perception of lexical tones as early as 4 months: English-, Cantonese-, and Mandarin-exposed infants each demonstrated different discrimination abilities that accorded with the properties of their native language at this stage. This study suggests that the formation of tone categories evolves earlier than that for vowels and consonants. What was previously regarded as a language-general stage of phonological development (from birth to 6 months) appears not to be so for infants with tone languages as their mother tongue.

Modern Mandarin is a tone language with relatively simple syllable structure (see Figure 1). Each syllable must be attached with one of the four lexical tones distinguishing different meanings (Tone 1: high-level tone; Tone 2: mid-rising tone; Tone 3: low-falling-rising tone; and Tone 4: high-falling tone). Lexical tones and nuclear vowels are compulsory elements for Mandarin syllables, whereas the onsets and the ending consonants are optional (Wang, 1973). Hua and Dodd (2000) systematically studied the production of Mandarin among one- to four-year-old children, reporting that children acquired phonological elements in the following order: tones were acquired first, followed by vowels and syllable-final consonants, which were then followed by syllable-initial consonants. The phonological saliency hypothesis (Hua & Dodd, 2000) might account for the order of phonological production in Mandarin, with Mandarin tones being the most salient. Moreover, in a large-scale population-based production study with children aged two to twelve years (To, Cheung & McLeod, 2013), results also showed that tone production was early acquired, followed by vowels and then by consonants.

Earlier studies (e.g. Chao, 1951; Clumeck, 1980; Hua, 2002; Hua & Dodd, 2000; Li & Tompson, 1977) have indicated that lexical tones are produced early, with considerable accuracy before age three. Judgments of tone errors in these studies were typically made by native adult observers in a categorical fashion (correct or incorrect). However, it is important to note

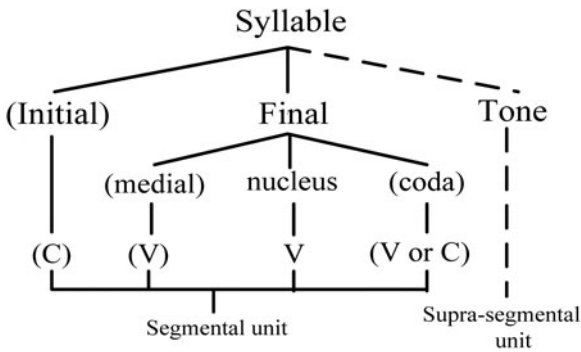

Fig. 1. Diagram of modern Mandarin syllable (elements in the parentheses are optional).

that the aforementioned early production of tones in children before three years of age does not mean that these children have the same abilities of tonal production as do adults. Innovations in speech analysis tools have afforded greater precision in the evaluation of early tone productions. A series of studies (Wong, 2012, 2013; Wong, Schwartz & Jenkins, 2005) have reported that three- to five-year-old preschoolers have not yet fully mastered the production of Mandarin tones. None of the Mandarin tones produced by the three- to five-year-old children reached adult-like accuracy, suggesting a protracted course of development extending beyond age five. Moreover, Yang, Diehl, and Davis (2008) investigated the duration of Mandarin tones produced by five-, eight-, and twelve-year-old monolingual Mandarin-speaking children and young adults. Although five- and eight-year-olds have already established lexical contrasts of Mandarin tones, adult-like tone production is still developing, with tone duration approximating adult values only in twelve-year-old children. These findings stood in contrast to earlier studies that claim very early acquisition of stable tone productions (Chao, 1951; Clumeck, 1980; Hua, 2002; Hua & Dodd, 2000; Li & Tompson, 1977). Furthermore, Wong's studies suggested that tones did not necessarily mature in advance of vowels and consonants in production (Wong, 2012, 2013).

For Mandarin tone perception, although the study of three-year-old children suggests they already show relatively high perceptual accuracy (around 90%) of all four Mandarin tones (Wong *et al.*, 2005), research on the developmental course of categorical perception (CP) of Mandarin tones is still lacking. The study of CP is useful because it offers a much more refined perceptual method to track the course of maturation and stabilization in children's fine-grained perception of Mandarin tones beyond age three.

CP is one of the most extensively studied phenomena in speech perception over the past fifty years. It is important in cognitive science because it reflects

the adaptation of the perceptual system to facilitate categorization among complicated variants that an organism needs to make in its environment (Goldstone & Hendrickson, 2010). In the auditory modality of speech perception, CP refers to the fact that listeners rapidly map continuous acoustic signals onto discrete phonological categories, resulting in better discrimination of stimuli across the category boundary than of equivalently separated stimuli within the same category (Liberman, Harris, Hoffman & Griffith, 1957). While the great bulk of literature on CP deals with the segmental features (i.e. consonants and vowels) (e.g. Fry, Abramson, Eimas & Liberman, 1962; Harnad, 1987; Johnson & Ralston, 1994; Liberman, Harris, Kinney & Lane, 1961; Miller & Eimas, 1977; Nenonen, Shestakova, Huotilainen & Näätänen, 2003), there has been a recent surge of research into whether CP applies to the suprasegmental level, such as lexical tone contrasts in tone languages (e.g. Bent, Bradlow & Wright, 2006; Bidelman, Gandour & Krishnan, 2011; Francis, Ciocca & Ng, 2003; Gandour, Wong & Hutchins, 1998; Hallé, Chang & Best, 2004; Peng, Deutsch, Henthorn, Su & Wang, 2013; Peng, Zheng, Gong, Yang, Kong & Wang, 2010; Wang, 1976; Xu, Gandour & Francis, 2006).

On the whole, studies on the CP of lexical tones in adults have mainly focused on two aspects: the influence of language experience (cross-language or cross-dialect), and different domains (music vs. speech vs. non-speech). Specifically, for Mandarin tone perception, there is ample evidence that adult Mandarin listeners not only show a sharp boundary (or a steep transition slope) between the level tone and the rising tone in the identification function, but they also exhibit a prominent discrimination peak around the categorical boundary position, indicating a typical CP of these Mandarin tones (e.g. Peng *et al.*, 2010; Wang, 1976; Xu *et al.*, 2006).

Compared to the extensive research on CP in adults, the progressive development of CP of Mandarin tones from infancy into childhood has received relatively little attention (but see Xi, Jiang, Zhang & Shu, 2009; Yang & Liu, 2012; Zhang *et al.*, 2012). In a behavioral study by Xi *et al.* (2009), healthy five- to seven-year-old children and adults were tested on the identification of a tone continuum ranging from Mandarin Tone 1 to Tone 2. Their results showed that six-year-old first-graders have already acquired adult-like identification competence for these Mandarin tones. However, due to the lack of an explicit discrimination test, their results may not be conclusive. In another behavioral study (Yang & Liu, 2012), eight Mandarin monolingual children aged from six to eight years (mean = 7·2) participated in both tone identification and tone discrimination tasks for tone continua between Tone 1 and Tone 2, and between Tone 1 and Tone 4. Mandarin monolingual children showed sharp categorical boundaries in the tone identification task. However, they did not show prominent discrimination peaks around the observed identification

boundaries, implying that the discrimination ability of Mandarin tones in six- to eight-year-old children was not as proficient as it was in adults. Zhang *et al*. (2012) made use of both behavioral and electrophysiological measures to compare ten-year-old Chinese dyslexic children with age-matched healthy controls. Zhang *et al*.'s results demonstrated that the perception of Mandarin high-rising Tone 2 and falling Tone 4 was already categorical among the healthy ten-year-old Chinese children. All in all, the general course of the development of CP of Mandarin tones in young children, who are still developing their ability for Mandarin tone perception, is not yet clear.

Behavioral tests for CP often require overt responses and sustained concentrations, which are relatively difficult for young children. This may be one of the reasons why behavioral research of CP in children is often limited and inconclusive. In the current study, a picture-pointing method was adopted to study young children's responses in both the identification and the discrimination tasks (see 'Procedure').

Furthermore, in Mainland China, children normally attend the first grade of primary school at the age of six or seven, where they begin to receive systematic training in phonological coding (i.e. *Pinyin*), including tone instruction. In other words, under normal conditions, six-year-olds in kindergarten have little metalinguistic knowledge of Mandarin tones, while children of the same age who happen to be in the first grade typically do have knowledge of the tone concept. In the study by Xi *et al*. (2009), six-year-olds in first grade showed a significant enhancement in the tone identification task, which could be potentially attributed to two factors: physiological maturation at the age of six and/or explicit instruction of Mandarin tone categories in first grade. However, which of these factors (if either) plays the more important role for the identification improvement has not yet been determined. To shed light on this issue, in the current study, young children were divided into five groups: four-, five-, and six-year-olds in kindergarten, and six- and seven-year-olds in first grade. Additionally, a control group of Mandarin-speaking young adults was recruited. It is important to investigate the developmental trajectory of Mandarin tone perception in the context of physiological growth and tone education, and, in particular, the earliest time-point at which children typically begin to perceive Mandarin tones categorically in the same way as adults, in both identification and discrimination tasks.

METHOD

Participants

Seventy-five young children aged four to seven years with Mandarin as their native language were recruited for this study. However, two from the

four-year-old group and one from the five-year-old group failed to meet the criteria on the training trials. Moreover, two children, one each from the four-year-old group and the seven-year-old group, were reluctant to cooperate during the process of testing and failed to complete the tasks. All the above-mentioned five child subjects were excluded and are not further shown in the current study. Additionally, a control group was recruited that consisted of sixteen young adults (8 females; mean age = 26·2, $SD = 2·3$ years) from the northern part of China. All participants were free of self-reported neurological diseases, psychiatric disorders, or hearing deficits. Informed consent (using a form approved by the Behavioral Research Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences) was obtained from all parents and children. The children were recruited from two kindergartens and one primary school, which had students from roughly equivalent family backgrounds in terms of both socioeconomic status and the parents' education levels. Of these, 14 children were in the four-year-old group from kindergarten (7 girls), 14 were in the five-year-old group from kindergarten (9 girls), and 14 were in the six-year-old group from kindergarten (8 girls). Another 14 children were also in the six-year-old group but from a primary school in first grade (6 girls), and 14 were in the seven-year-old group also in first grade (6 girls). We refer to these five groups as: four yr, five yr, six yr (K), six yr (P), and seven yr throughout this paper. The non-verbal intelligence of the children had been assessed by a local administrant hospital before children went to school, using Raven's Standard Progressive Matrices (Raven, 1976). All child participants showed normal non-verbal intelligence. Moreover, since Gathercole and Baddeley (1990) argued that children with specific language impairments (SLI) have reduced phonological storage in their working memory, every child's language ability (LA) was screened with the Test of Language Ability in Preschoolers (Ning, 2013), which consists of five subtests (including grammatical competence, basic vocabulary, and semantic application). Results indicated all child participants showed normal language development. The basic information of the tested children is shown in Table 1.

Materials

The syllable /i/ with high-level tone (around 290 Hz) produced by a female native Mandarin speaker was recorded (22050 Hz sampling rate, 16-bit resolution). This female speaker was chosen because her F0 range was close to that of the average female voice (Peng, Zhang, Zheng, Minett & Wang, 2012). Based on the natural speech template with Tone 1, all tone stimuli were re-synthesized by applying the pitch-synchronous overlap and

TABLE 1. *Characteristics of children in five age groups*

	Four yr	Five yr	Six yr (K)	Six yr (P)	Seven yr
Number	14	14	14	14	14
Age (year; month)	4;1–4;11	5;2–5;10	6;1–6;10	6;2–6;11	7;1–7;10
Mean age (year)	4·52 ± 0·24	5·53 ± 0·21	6·24 ± 0·19	6·4 ± 0·23	7·25 ± 0·21
Accuracy of LA (%)	91·24 ± 3·88	92·58 ± 4·54	95·58 ± 3·21	95·28 ± 3·6	95·86 ± 2·98

add method (Moulines & Laroche, 1995) implemented in Praat (Boersma & Weenink, 2009). Since the pitch pattern (i.e. fundamental frequency as its acoustic correlate) plays the primary role in tone perception in Mandarin (Kuo, Rosen & Faulkner, 2008), other acoustic cues which vary greatly in natural speech (e.g. amplitude, envelope, and duration) were kept constant among stimuli. The procedures for synthesizing the stimuli followed those described in Peng *et al*. (2010). Figure 2 shows a schematic diagram of the pitch contours of the eleven stimuli along the tone continuum (following Peng *et al*., 2010; Wang, 1976). The F0 onsets of the target stimuli were determined by the formula '230 Hz + 6 Hz × (Stimulus Number − 1)'. In Mandarin, the syllable /i/, when produced with the high-level tone, is a homophone with fifteen different meanings such as 'clothes', and is coded as the typical stimulus #11; when /i/ is spoken with the mid-rising tone, it is another homophone with twenty-four different meanings such as 'aunt', which is coded as the typical stimulus #1. All the 500-ms tone stimuli were presented binaurally at 70 dB sound pressure level.

Procedure

To help ensure young children would be able to accomplish the whole identification task and discrimination task (lasting 30 minutes and 50 minutes, respectively), the experimental designs were modified to make these tasks more manageable for our child participants, and the two tasks were separately conducted on two successive days.

The identification task was conducted on the first day. Before the experiment, a training session was provided to ensure that all the children could follow the instructions. The experimenter taught them to point at the left picture on a computer screen (a car driving on a level road) after playing the stimulus #11 several times, and to point at the right picture (a car driving on a rising road) after playing the stimulus #1 several times. Only after the children succeeded in pointing at the matched picture each time they heard stimulus #11 or #1 would they progress to the next step. There was one practice block (including stimuli #1, #2, #5, #6, #7, #10, #11, repeating twice randomly) before the two test blocks to familiarize children with the identification procedure. An identification accuracy of

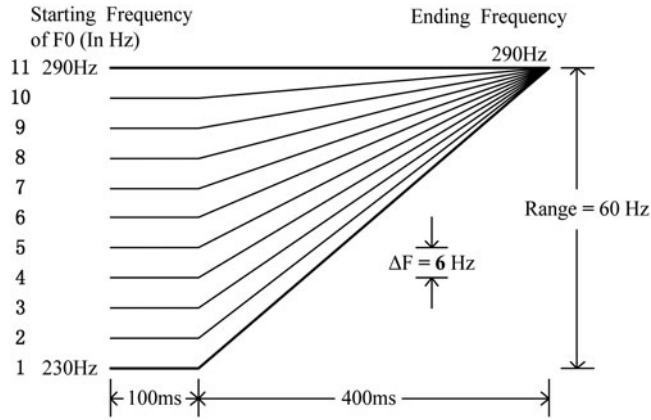

Fig. 2. Schematic diagram of the pitch contours.

more than 90% for the practice stimuli close to the two ends of the continuum (i.e. #1, #2, #10, #11) was required to continue to the test blocks. The eleven stimuli were randomly presented four times each in each test block. There were two test blocks for a total of eighty-eight test trials. The entire practice and test blocks were controlled by E-prime. The experimenter logged the child's pointing responses by pressing key '1' (for 'Tone 1') on the keyboard when children pointed at the left picture and key '2' (for 'Tone 2') when children pointed at the right picture (two-alternative forced choice, 2AFC). No feedback was given to the participants during the practice and test blocks.

The discrimination task was conducted on the following day. During the training session, the experimenter instructed them to point at the left picture (a happy face with two identical eyes) representing the same pairs after playing sound pairs (11–11 or 1–1) several times, and to point at the right picture (a sad face with two different eyes) representing the different pairs after playing the sound pairs (11–1 or 1–11) several times. Only after the children succeeded in pointing at the matched picture each time they heard the training sound pairs (11–11; 1–1; 11–1; 1–11) would they progress to the experiment. The practice block contained twelve pairs (5–7, 6–8, 7–5, 8–6, 1–1, 11–11, repeating twice randomly). No discrimination accuracy was requested for the practice block. For the testing part, twenty-nine pairs were presented in a random order, with a 500 ms inter-stimulus interval (ISI). Among the twenty-nine pairs, eighteen pairs consisted of two different stimuli (different pairs) separated by two steps (i.e. 12 Hz), in either forward (1–3, 2–4, 3–5 ... 8–10, 9–11) or reverse order (3–1, 4–2, 5–3 ... 10–8, 11–9), and eleven pairs consisted of all the eleven stimuli on the continuum each paired with itself (same

pairs). The above twenty-nine pairs were repeated five times, resulting in 145 testing pairs in total. The experimenter pressed 'v' (for 'same') on the keyboard when children pointed at the smiling picture and pressed 'n' (for 'different') when children pointed at the sad picture. The next sound pair would be presented 500 ms later, after the experimenter pressed the space bar, and children were free to have a rest whenever they wanted. No feedback was given to the participants during the practice and test blocks.

For the adult participants, the training session was excluded, and the materials in the practice and test blocks were exactly the same as for the child participants. The identification and discrimination tasks were conducted within one day. During the identification task, the adult participants were asked to press key '1' on the keyboard when they heard a 'Tone 1' sound and to press key '2' when they heard a 'Tone 2' sound. In the discrimination task, they were instructed to judge whether the two stimuli were the same or different, and to respond by pressing a key ('v' for 'same', and 'n' for 'different').

Data analyses

To investigate the developmental trajectory of the identification and discrimination performance, three key parameters of CP were calculated: position of category boundary, width of category boundary, and discrimination accuracy.

For each stimulus, the identification score was computed as the average percentage of level (Tone 1) or rising tone (Tone 2) responses. Boundary position and boundary width were assessed using Probit analyses (Finney, 1971). The boundary position (defined as the 50% cross-over point) and the boundary width (defined as the linear distance between the 25th and 75th percentiles) were analyzed according to the procedures described in Peng *et al.* (2010).

To calculate the discrimination scores, we divided the 145 discrimination trials into nine comparison units, each consisting of all pairs in four types of pairwise comparisons (AB, BA, AA, and BB) (cf. Xu *et al.*, 2006). Adjacent comparison units contained the overlapping AA or BB pairs (e.g. the five 4–4 pairs were included in both 2–4 and 4–6 units). The discrimination accuracy (P) for each comparison unit was calculated with the formula described in Xu *et al.* (2006):

$$P = P(\text{'S'}/S) \times P(S) + P(\text{'D'}/D) \times P(D)$$

where the percentages of 'same' ('S') and 'different' ('D') responses to all the same (S) and different (D) trials (i.e. the correct responses) were represented by P ('S'/S) and P ('D'/D), respectively. P(S) and P(D) were the probabilities of S (AA or BB) and D (AB or BA) trials in each unit, which were both equal to 0·5 (Peng *et al.*, 2010).

In addition, we calculated the discrimination accuracy of between-category comparisons (Pbc) and within-category comparisons (Pwc) for each subject based on the position of category boundary as follows (cf. Jiang, Hamm, Lim, Kirk & Yang, 2012). If the identification results indicated that one participant's categorical boundary was at position 5·5, then the average discrimination scores corresponding to the comparison units 4–6 and 5–7 which straddled the boundary would be coded as the discrimination accuracy for between-category comparisons, while the within-category discrimination accuracy for that subject was obtained from the mean of the discrimination scores for the remaining seven comparison units.

RESULTS

Identification and discrimination curves

Identification and discrimination curves among different age groups are shown in Figure 3. Similar to adult Mandarin listeners, categorical boundaries and corresponding discrimination peaks were found for all child participants, indicating a typical CP of Mandarin Tone 1 and Tone 2.

Position and width of categorical boundary

The estimated boundary position and boundary width values are shown in Table 2. In addition, the distribution of boundary width among different groups is shown in Figure 4.

The mean boundary positions for the groups are as follows: four yr: 5·81, five yr: 5·63, six yr (K): 5·66, six yr (P): 5·95, seven yr: 5·57, adults: 5·90. Results of one-way ANOVA revealed that the perceptual boundary positions were not significantly different among different age groups [$F(5,80) = 0{\cdot}638$; $p = {\cdot}671$]. The average boundary widths for the groups are as follows: four yr: 2·03, five yr: 1·93, six yr (K): 1·22, six yr (P): 1·09, seven yr: 1·02, and adults: 1·01. Three outliers (shown in Figure 4) were removed from the statistical analyses, with one in six yr (P) and two in seven yr. One-way ANOVA showed that the identification widths among different age groups were significantly different [$F(5,80) = 11{\cdot}148$; $p < {\cdot}001$]. Tukey's HSD post-hoc pairwise comparisons of the six groups indicated that the groups of six yr (K), six yr (P), seven yr, and adults had significantly narrower boundary widths than the four-year-old group (all ps $< {\cdot}001$) and five-year-old group (all ps $< {\cdot}01$). However, the boundary widths between the four- and five-year-old groups were not significantly different ($p = {\cdot}995$), and the boundary widths between any two groups of six yr (K), six yr (P), seven yr, and adults were not significantly different either (all ps $> {\cdot}8$).

To investigate the correlation between boundary widths and chronological ages, we conducted a Pearson correlation analysis and found a negative correlation between boundary widths and chronological ages ($r = -0{\cdot}465$,

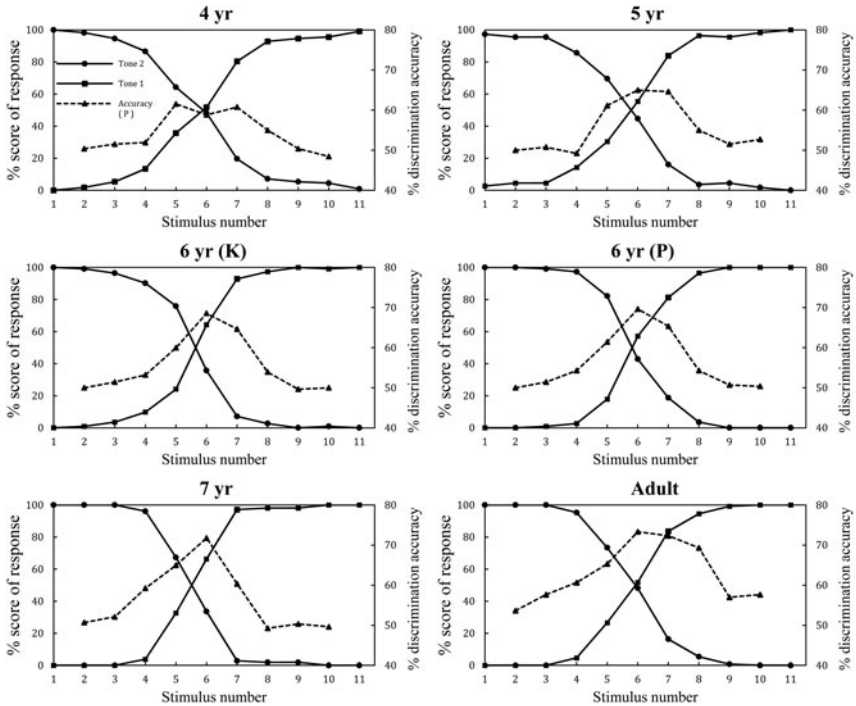

Fig. 3. The identification curves (solid lines, with round dots representing Tone 2 response and square dots representing Tone 1 response) and discrimination curves (dashed lines with triangle dots). The left y-axis indicates the percentage score of Tone 2 or Tone 1 response, while the right y-axis indicates the overall discrimination accuracy.

$p < \cdot 01$). We have further divided the child participants into two subgroups: one group contained four yr and five yr, the other group contained six yr (K), six yr (P) and seven yr. However, for the 28 four- to five-year-old children, Pearson correlation analysis indicated that there was no correlation between the boundary widths and ages ($r = -0{\cdot}123$, $p = \cdot 531$), and for the 42 six- to seven-year-old children, no significant correlation between the boundary widths and ages was found either ($r = -0{\cdot}180$, $p = \cdot 253$). Results indicated that the development of identification acuity was not continuous as a function of chronological age for children aged four to five, and for those aged six to seven.

Discrimination accuracy

The overall discrimination accuracies of the nine comparison units among the six groups are shown in Figure 5. For the four-year-old children, the

TABLE 2. *Derived categorical boundary position and width for each group*

	Position		Width	
Group	Mean	SD	Mean	SD
Four yr	5·81	0·98	2·03	0·69
Five yr	5·63	0·59	1·93	0·80
Six yr (K)	5·66	0·56	1·22	0·31
Six yr (P)	5·95	0·71	1·09	0·28
Seven yr	5·57	0·56	1·02	0·29
Adult	5·90	0·93	1·01	0·28

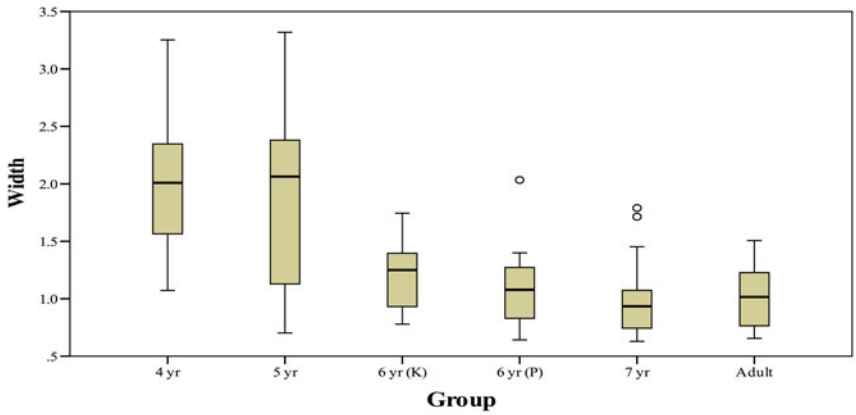

Fig. 4. Box plots of boundary widths within each age group, where the bold black line inside the boxes marks the median, and the upper and lower boundaries of the box correspond to its upper and lower quartiles.

discrimination peak was 61·54% at the comparison unit 4–6 straddling the boundary position at 5·81. For the children from groups of five yr, six yr (K), six yr (P), and seven yr, the discrimination accuracies all reached their maxima at the unit 5–7 (straddling the boundary position 5·57–5·95), which were equal to 65·00%, 68·57%, 69·94%, 71·79%, and 73·33%, respectively.

A two-way 9 (comparison unit) × 6 (age group) analysis of variance (ANOVA) was conducted to examine whether the six age groups showed different overall discrimination accuracy, with comparison unit as the within-subject factor and the age group as the between-subject factor. Where appropriate, the Greenhouse–Geisser adjustment method was used to correct violations of sphericity. The analysis confirmed a significant main effect for comparison unit [$F(8,616) = 60{\cdot}068$; $p < {\cdot}001$], and group

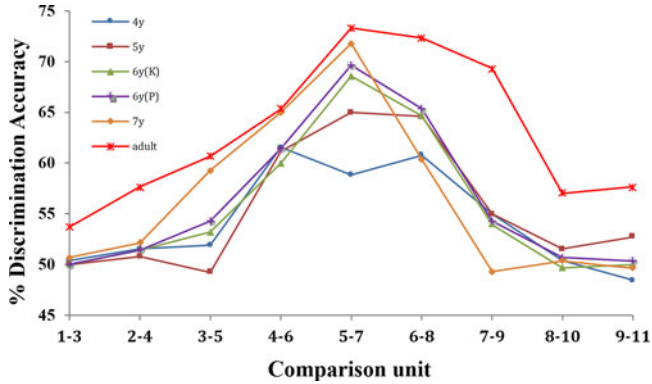

Fig. 5. The discrimination accuracy of nine comparison units across six groups.

[$F(5,77) = 11{\cdot}091$; $p < {\cdot}001$]. There was a significant interaction between comparison unit and group [$F(40,616) = 1{\cdot}793$; $p < {\cdot}05$]. Post-hoc analysis indicated that the overall discrimination accuracy in the adult group was significantly higher than the child groups (all ps $< {\cdot}001$), while the overall discrimination accuracies between any two of the five child groups were not significantly different (all ps $> {\cdot}05$).

The overall discrimination accuracies of nine comparison units were further divided into between-category comparison (Pbc) and within-category comparison (Pwc) as shown in Figure 6. The average between-category discrimination accuracy (Pbc) for different age groups are as follows: four yr: 60·54%, five yr: 62·32%, six yr (K): 65·00%, six yr (P): 65·71%, seven yr: 66·96%, and adults: 71·33%. One-way ANOVA showed that the between-category discrimination accuracy among different age groups were significantly different [$F(5,79) = 2{\cdot}437$; $p < {\cdot}05$]. However, Tukey's HSD post-hoc pairwise comparisons showed that the between-category discrimination accuracy was not different between any adjacent two child groups (all ps $> {\cdot}05$). Moreover, the mean within-category discrimination accuracy (Pwc) for the groups of four yr, five yr, six yr (K), six yr (P), seven yr, and adults were 52·55%, 53·57%, 53·06%, 53·72%, 53·52%, and 60·62%, respectively. Results of a one-way ANOVA revealed that the within-category discrimination accuracy was also significantly different among different age groups [$F(5,79) = 12{\cdot}443$; $p < {\cdot}001$]. However, Tukey's HSD post-hoc pairwise comparisons showed that the within-category discrimination accuracy was similar between any two adjacent child groups (all ps $> {\cdot}05$).

To investigate the correlation between between-category discrimination accuracies (Pbc) and chronological ages, a Pearson correlation analysis was further conducted between Pbc and chronological ages, and the

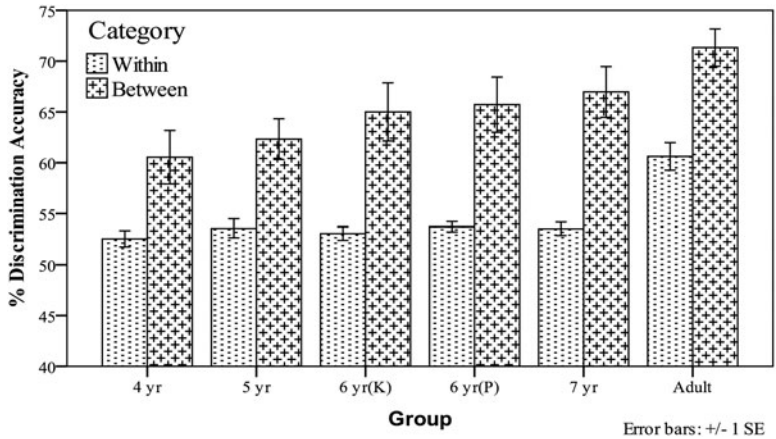

Fig. 6. Between-category and within-category discrimination accuracy pooled across different age groups.

corresponding scatterplot is shown in Figure 7. A positive correlation between Pbc and chronological ages ($r = 0{\cdot}238$, $p < {\cdot}05$) was found. However, for the within-category discrimination accuracy (Pwc), there was no significant correlation between within-category discrimination accuracies and ages ($r = 0{\cdot}098$, $p = {\cdot}418$).

In conclusion, from the identification curves and the distribution of boundary widths, we could see clearly that changes from high-level to mid-rising tone were more abrupt from six-year-olds. For the tone discrimination, the sensitivity to between-category discrimination showed a rising trend as children get older. However, the within-category discrimination accuracies did not change significantly among all the child participants. Moreover, children as early as four years of age, with plenty of exposure to Mandarin as their mother tongue, could perceive Mandarin Tone 1 and Tone 2 categorically.

DISCUSSION

The current study examined CP of Mandarin Tone 1 and Tone 2 in typically developing four- to seven-year-old children. As defined by Liberman *et al.* (1957), typical CP should be signaled by a sharp boundary between two categories and a corresponding discrimination peak straddling the identification boundary. In this study, prominent discrimination peaks were well aligned with the corresponding identification cross-overs for all child participants, indicating that tone perception was highly categorical (see Figure 3). These Mandarin-speaking young children have formed their perceptual ability to process lexical tones as different categories, even

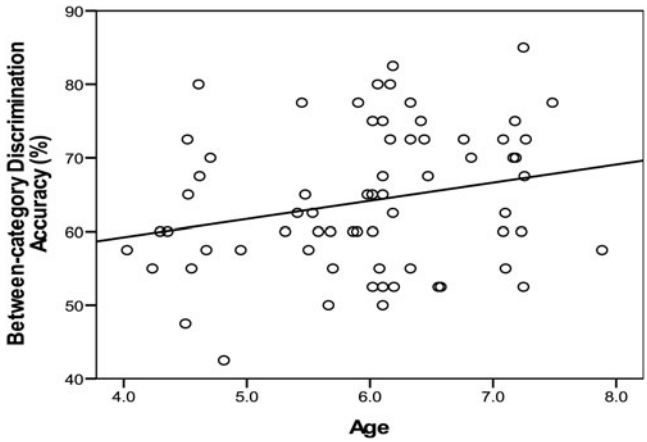

Fig. 7. Correlations between chronological age and between-category discrimination accuracy.

without knowing the metalinguistic concept of tones at all. In stark contrast, six- to eight-year-old English monolingual children who have not yet been exposed to a tone language showed no categorical boundaries or discrimination peaks for the continuum between Mandarin Tone 1 and Tone 2 (Yang & Liu, 2012). Therefore, it is reasonable to conclude that tone language exposure is critical for learning to perceive lexical tones categorically.

Tone contrasts of different categories may emerge very early in perception, but the process actually takes a relatively long time to fully mature. The development of CP of Mandarin Tone 1 and Tone 2 in young children with respect to both identification and discrimination performance was observed in the current study. In the present identification test, different age groups did not show significantly different boundary positions (ranging from 5·57 to 5·95; see Table 2). Similarly, adults with tone or non-tone language experiences yielded similar identification boundaries near the middle of the tone continuum (e.g. Peng *et al.*, 2010; Xu *et al.*, 2006). These findings seem to decrease the efficacy of boundary position in assessing the degree of CP. However, the boundary width was much narrower (i.e. a sharper boundary or steeper transition slope) for tone language listeners than for non-tone language listeners (Hallé *et al.*, 2004; Peng *et al.*, 2010; Wang, 1976; Xu *et al.*, 2006). Likewise, the current experimental data showed that child groups of six yr, seven yr, and the adult group exhibited much narrower boundary widths than those of the four- and five-year-old groups (see Table 2 and Figure 4), indicating that six-year-olds have already acquired adult-like identification competence of

Mandarin high-level and mid-rising tones. The narrower boundary width from age six was due to higher identification acuity of the middle ambiguous stimuli, as reflected by a sharper boundary. Our conclusion is a supplement to another identification study by Xi *et al*. (2009), in which a significant enhancement in tone identification of the tone continuum between Tone 1 and Tone 2 was found for six-year-olds in first grade who had already received tone instruction. What we would like to emphasize here is that none of the four- to six-year-old preschoolers in our study had received explicit education concerning the concept of Mandarin tones, and, in the identification task, they were simply asked to judge the sound stimulus like a level sound or a rising one. Consequently, it is reasonable to conclude that the adult-like identification competence of Mandarin Tone 1 and Tone 2 appears around age six, regardless of explicit education in tone categories.

In addition to the identification improvement at age six, there was also a performance increment with age for the Mandarin-speaking children's between-category discriminability, but not for the within-category discriminability. These findings are consistent with the dual-process model (Fujisaki & Kawashima, 1971), which proposes two separate pathways for the discrimination of speech stimuli. Fujisaki and Kawashima assume first and foremost that identification is inevitable, even in discrimination experiments where identification is not a part of the subject's task. In this model, two separate pathways in the discrimination task are involved in the decision mechanism. One pathway is used when the two stimuli of the comparison pair receive different phoneme labels. The labeling process along this first pathway is based on the information of phoneme boundaries stored in long-term memory (i.e. phonetic memory code). If the two stimuli that have to be discriminated instead receive the same label, then, and only then, the other pathway is used, and the information about the spectral composition of the two stimuli (based on short-term auditory memory code) leads to the final decision. Both auditory and phonetic modes co-exist during the discrimination task, but subjects operate exclusively in long-term phonetic mode during the identification task. Consequently, according to this model, both the identification and between-category discrimination abilities undergo a developmental pattern when children are exposed to a tonal environment consistently, and further store the tone category information in their long-term phonetic memory.

On the other hand, from the perspective of Fujisaki and Kawashima's (1971) model, if children identify the two-tone stimuli as being in the same category, the second pathway then kicks in. Children would make their decision on the basis of the physical F0 difference between the two stimuli, which is based on short-term auditory memory code. Our results

revealed no significant improvement in the performance of within-category discrimination in four- to seven-year-old children, with each stimulus pair separated by 12 Hz at the onset F0. As suggested in a recent study (Liu, 2013) exploring the just noticeable difference (JND) of the F0 contour, Mandarin adult listeners required nearly 10 Hz to detect a tonal pitch change. Given that young children generally require greater JND in F0 discrimination than adult listeners (Elliot, Hammer, School & Wasowicz, 1989; Jensen & Neff, 1993), children may require a F0 difference of more than the 12 Hz used in the current study, even at age seven, to perform the within-category discrimination task more reliably.

In addition to the dual-process model, there is another approach called the 'signal-detection model' (Durlach & Braida, 1969), with two modes of memory operation included: a trace mode and a context-coding mode. Applying this signal-detection model to CP successfully explains the discrepancy between fixed/roving discrimination and identification sensitivity regarding differences in experimental context (Macmillan, Kaplan & Creelman, 1977). However, the primary drawbacks of the signal-detection model lie in that it does not differentiate short-term and long-term memory components in the context-coding mode (Macmillan, Goldberg & Braida, 1988), and this model only considers the different variants of the experimental context. Consequently, the signal-detection model could not explain any experience-driven effects on CP. For example, many cross-language CP studies (Hallé *et al.*, 2004; Peng *et al.*, 2010; Wang, 1976; Xu *et al.*, 2006) indicated that tonal language listeners demonstrated a much sharper identification boundary and a more prominent discrimination peak around the boundary than non-tonal language listeners. Cross-language differences can only be explained with regard to differences in the two groups' long-term experience with their respective native languages, and not due to differences in experimental context (that was identical for the tonal language and non-tonal language listeners). Similarly, in the present study, age-related enhancement of both identification and between-category discrimination ability in children can be attributed to long-term accumulation of tone language experience, and not to the differences in the experimental context that was identical for all the child participants. Therefore, the results of the present study fit in with previous cross-language research and provide empirical evidence for the dual-process model from the development of CP of tones in native Mandarin-speaking children.

As proposed by Wang (1978), the language that children use changes as a function of ambient language input. Therefore, we assumed that the perceptual progress of Mandarin tones could be driven by perceptual accumulation in the form of 'distributional learning'. Maye *et al.* (2002) familiarized infants with speech sounds from a phonetic continuum, exhibiting either a bimodal or a unimodal frequency distribution. During

the test phase, only infants exposed to the bimodal condition discriminated tokens from the endpoints of the continuum, suggesting that there might be a distributional learning mechanism through which infants track the regularities of the sounds in the ambient language. Moreover, sensitivity to the distributional patterns in the input is not restricted to infancy, but contributes to learning throughout one's lifetime (Maye, 2000; Saffran, 2001). While growing up, Mandarin-speaking children may be able to use the bimodal frequency distribution from the tonal information of the ambient sound input to develop the perceptual ability to identify and discriminate different tonal categories.

Consequently, the development of CP from infancy into childhood is presumably influenced by accumulated exposure to a tone language, which causes the between-category differences to become more salient, thus enhancing the discriminability of pitch variations across different tone categories with different lexical meanings. Moreover, during the maturation and stabilization of tone categories, the identification acuity around the category boundary of Mandarin Tone 1 and Tone 2 reached adult-like performance from age six, facilitating Mandarin-speaking children to identify different tone categories around boundary more confidently, so they could better master the many-to-one mapping between acoustic patterns and phonological tone categories, and accordingly enhance spoken communication.

CONCLUSION

Using the classic paradigm of CP, the current study sought to explore the development of Mandarin tone perception in four- to seven-year-old children. Mandarin-speaking children, with exposure to native tone language, can perceive Mandarin Tone 1 and Tone 2 categorically. The positions of the identification boundaries did not differ significantly between children and adults, but the boundary widths between Tone 1 and Tone 2 did differ significantly, with much narrower boundary widths (i.e. sharper boundaries) occurring in six-year-olds than five-year-olds. Moreover, with age, the ability to discriminate more fine-grained tonal differences of between-category pairs was enhanced gradually due to perceptual accumulation. These findings contribute to the work in discovering the general developmental course of CP during the maturation and stabilization of tone perception in young children.

REFERENCES

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**, 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**, 183–216.

Best, C. T., McRoberts, G. W., LaFleur, R. & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behavior and Development* **18**, 339–50.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second language speech perception: commonalities and complementarities. In O.-S. Bohn & M. J. Munro (eds), *Language experience in second language speech learning: in honor of James Emil Flege*, 13–34. Amsterdam: John Benjamins.

Bidelman, G. M., Gandour, J. T. & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience* **23**, 425–34.

Boersma, P. & Weenink, D. (2009). Praat: Doing phonetics by computer. Online: <http://www.fon.hum.uva.nl/praat/>.

Chao, Y. R. (1951). The Cantian idiolect: an analysis of the Chinese spoken by a twenty-eight-months-old child. In W. J. Fischel (ed.), *Semitic and oriental studies* (pp. 27–44). Berkeley, CA: University of California Press.

Clumeck, H. (1980). The acquisition of tone. In G. H. Yeni-Komshian, J. F. Kavanaugh & C. A. Ferguson (eds), *Child phonology: Vol. 1. Production*, 257–75. New York: Academic Press.

Durlach, N. I. & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America* **46**, 372–83.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W. & Vigorito, J. (1971). Speech perception in infants. *Science* **171**, 303–6.

Elliot, L. L., Hammer, M. A., School, M. E. & Wasowicz, J. M. (1989). Age differences in discrimination of simulated single-formant frequency transitions. *Perception and Psychophysics* **46**, 181–6.

Finney, D. J. (1971). *Probit analysis*, 3rd ed. Cambridge: Cambridge University Press.

Francis, A. L., Ciocca, V. & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics* **65**, 1029–44.

Fromkin, V. (1978). *Tone: a linguistic survey*. New York: Academic Press.

Fry, D. B., Abramson, A. S., Eimas, P. D. & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech* **5**, 171–89.

Fujisaki, H. & Kawashima, T. (1971). A model of the mechanisms for speech perception-quantitative analysis of categorical effects in discrimination. Annual Report of the Engineering Research Institute, Faculty of Engineering, *University of Tokyo* **30**, 59–68.

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

Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language* **29**, 336–60.

Goldstone, R. L. & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science* **1**(1), 69–78.

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**, 395–421.

Harnad, S. (ed.) (1987). *Categorical perception: the groundwork of cognition*. New York, NY: Cambridge University Press.

Hua, Z. (2002). *Phonological development in specific contexts: studies of Chinese-speaking children*. Clevedon: Multilingual Matters.

Hua, Z. & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* **27**, 3–42.

Jensen, J. K. & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science* **4**, 104–7.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J. & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition* **40**, 1109–21.

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

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience* **5**, 831–43.

Kuhl, P. K. (2009). Early language acquisition: neural substrates and theoretical models. In M. S. Gazzaniga (ed.), *The Cognitive Neurosciences IV*, 837–54. Cambridge, MA: MIT Press.

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**, F13–21.

Kuo, Y. C., Rosen, S. & Faulkner, A. (2008). Acoustic cues to tonal contrasts in Mandarin: implications for cochlear implants. *Journal of the Acoustical Society of America* **123**, 2815–24.

Lenneberg, E. H. (1967). *The biological foundations of language*. New York: Wiley.

Li, C. N. & Tompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* **4**, 185–99.

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**, 358–68.

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**, 379–88.

Liu, C. (2013). Just noticeable difference of tonal pitch changes in speech and nonspeech stimuli for Chinese- and English-native speakers. *Journal of the Acoustical Society of America* **134**, 3011–20.

Macmillan, N. A., Goldberg, R. F. & Braida, L. D. (1988). Resolution for speech sounds: basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America* **84**, 1262–80.

Macmillan, N. A., Kaplan, H. L. & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review* **84**, 452–71.

Mattock, K. & Burnham, D. (2006). Chinese and English infants' tone perception: evidence for perceptual reorganization. *Infancy* **10**, 241–65.

Maye, J. (2000). *Learning speech sound categories on the basis of distributional information*. Unpublished doctoral dissertation, University of Arizona, Tucson.

Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, B101–11.

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. *Journal of the Acoustical Society of America* **61**, 835–45.

Moulines, E. & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *S*peech Communication **16**, 175–205.

Nenonen, S., Shestakova, A., Huotilainen, M. & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research* **16**, 492–5.

Ning, C. Y. (2013). *Test of Language Ability in Preschoolers*. Institute of Linguistics, Tianjin Normal University: Tianjin University Press [in Chinese].

Peng, G., Deutsch, D., Henthorn, T., Su, D.-J. & Wang, W. S.-Y. (2013). Language experience influences non-linguistic pitch perception. *Journal of Chinese Linguistics* **39**, 447–67.

Peng, G., Zhang, C. C., Zheng, H. Y., Minett, J. M. & Wang, W. S.-Y. (2012). The effect of inter-talker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research* **55**, 579–95.

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* **38**, 616–24.

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* **20**, 421–35.

Raven, J. C. (1976). *Standard progressive matrices*. Oxford: Oxford Psychologists Press.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language* **44**, 493–515.

Saffran, J. R., Werker, J. F. & Werner, L. A. (2006). The infant's auditory world: hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (eds), *Handbook of child psychology: Volume 2, Cognition, perception and language VI*, 58–108. New York: Wiley.

Streeter, L. A. (1976). Language experience of two-month-old infants shows effects of both innate mechanisms and experience. *Nature* **259**, 39–41.

To, C. K. S., Cheung, P. S. P. & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, and Hearing Research* **56**, 103–22.

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

Wang, W. S.-Y. (1973). The Chinese language. *Scientific American* **228**, 50–63.

Wang, W. S.-Y. (1976). Language change. In S. R. Harnad (ed.), *Origins and evolution of language and speech*, Vol. 280, 61–72. New York, NY: New York Academy of Sciences.

Wang, W. S.-Y. (1978). The three scales of diachrony. In B. B. Kachru (ed.), *Linguistics in the seventies*, 63–75. Urbana, IL: Department of Linguistics, University of Illinois.

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

Wong, P. S. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* **40**, 141–51.

Wong, P. S. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *Journal of the Acoustical Society of America* **133**, 434–43.

Wong, P., Schwartz, R. G. & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research* **48**, 1065–79.

Xi, J., Jiang, W., Zhang, L J. & Shu, H. (2009). Categorical perception of VOT and lexical tones in Chinese and the developmental course. *Acta Psychologia Sinica* **41**, 572–9.

Xu, Y., Gandour, J. T. & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America* **120**, 1063–74.

Yang, J., Diehl, R. & Davis, B. L. (2008). The development of tonal duration in Mandarin-speaking children. *Journal of the Acoustical Society of America* **124**, 2498 (A).

Yang, J. & Liu, C. (2012). Categorical perception of lexical tone in 6- to 8-year-old monolingual and bilingual children. *International Journal of Asian Language Processing* **22**, 49–62.

Yeung, H. H., Chen, K. H. & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language* **68**, 123–39.

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

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y. & Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry* **53**, 874–82.