Categorical Perception of Lexical Tone in 6 to 8-year-old Monolingual and Bilingual Children

Jie Yang and Chang Liu

Department of Communication Sciences and Disorders
The University of Texas at Austin, 1 University Station A1100, Austin, TX 78712
thyjessie@utexas.edu, changliu@mail.utexas.edu

ABSTRACT

Given that Mandarin is a tone language and English is not, the present study investigated the perception of Mandarin tones in young school-aged children with or without tone language exposure. Mandarin Monolingual (MM), English-Mandarin Bilingual (EMB) and English Monolingual (EM) children aged 6 to 8 years participated in tone identification and discrimination tasks. Tone identification and discrimination scores of manipulated Tone 1-Tone 2 and Tone 1-Tone 4 fundamental frequency continuum were measured. MM and EMB children exhibited categorical boundaries in tone identification, but EM children did not. MM children showed steeper identification slopes than EMB children. Tone discrimination showed modest peaks across tone categories for the MM and EMB children, but did not present peaks for the EM children, possibly due to underdeveloped capacity to discriminate tonal glides for children. Performance differences across the three groups could be possibly attributed to children's varied language backgrounds. Tone language exposure seemed critical for children to perceive lexical-level tonal changes.

Keywords

Lexical tone perception; categorical perception; Chinese and English children.

1 Introduction

In tone languages, besides other acoustic cues like syllable duration and intensity contour, fundamental frequency F_0 (height and/or contours) is used as a primary cue to contrast lexical meanings (Gandour & Harshman, 1978; Gao, 2002). For example, in Mandarin, the syllable [ma] with four different F_0 contour shapes (high-level as tone 1, high-rising as tone 2, low-fall-rise as tone 3, and high-falling as tone 4) means "*mother*", "*hemp*", "*horse*", "*scold*" respectively (Chao, 1968). That is, in general, a flat F_0 pattern is associated with tone 1 and rising and falling F_0 contours were associated with tone 2 and 4, respectively, while a falling-then-rising F_0 contour refers to tone 3. F_0 also provides intonation information at the utterance level in tone languages (Yuan, 2004). Perception of lexical tones is critical to understanding word meanings for effective linguistic communication in tone languages. In non-tone languages, F_0 is used only at the utterance level to encode intonation (Lehiste, 1970). Non-tone language listeners need to distinguish F_0 variations within larger linguistic units and may be less attuned to F_0 variations at the syllabic level (Gandour & Harshman, 1978).

They have shown significant difficulties in perceiving lexical tones (Kiriloff, 1969; Bluhme & Burr, 1971; Wang, Spence, Jongman, & Sereno 1999; Huang, 2004).

Tonal perception studies with infants have found that discrimination performance of tone and non-tone language listeners started to diverge as early as 6-9 months, when tonal perception was reorganized as a result of ambient language exposure (Mattock & Burnham, 2006). Investigations of speech output patterns indicated using F_0 as a cue to carry lexical and/or intonational information continued to develop through the school years, parallel to other aspects of speech production with a prolonged course of development (see Kent & Vorperian, 1995 for an overview). Children's ability to produce intonation functionally in non-tone languages began to be mastered by age 4-5 and stabilized with the maturation of their motor system capacities (Wells, Peppé, & Goulandris, 2004; Patel & Grigos, 2006; Patel & Brayton, 2009). Seven-year-olds employed variable acoustic cues, e.g. duration, intensity, and F_0, for the purposes of intonation contrast, but 11-year-olds mainly depended on adult-like F_0 cues in production output. Yang, Diehl, and Davis (2008) reported that the duration of lexical tone decreased to approach adult value by 8 years in Mandarin children despite the earlier establishment of phonological contrasts of tone. Patel and Brayton (2009) suggested that variability in speech prosodic production might result from English-native children's prosodic perception strategies which could be different from adults. Altogether, children appeared to develop their strategy of using different acoustic cues (e.g., F_0 contour and duration) to differentiate intonations and/or lexical tones in their speech production up to 8-to-11 years old (Patel & Grigos, 2006; Yang et al., 2008; Patel & Brayton, 2009). On the other hand, it is also important to understand how young children perceptually recognize F_0 contours of speech sounds, when they are more attuned to language-specific uses of F_0 compared to infants. However, perception of F_0 variations at either the syllable or utterance level has not been examined cross-linguistically across such extended periods of development. Thus, the present study was to evaluate lexical tone perception for school-aged children of 6-to-8 years old with and without tone language backgrounds.

One frequently investigated question in lexical tone perception is whether tone is perceived categorically similar to stop consonants (e.g. Liberman, Harris, Hoffman, & Griffith, 1957), or non-categorically resembling vowels (Fry, Abramson, Eimas, & Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969) as has been shown for non-tone language listeners. Studies of tonal perception in adult tone language listeners uniformly report category boundaries with sharp slopes in identification tasks in various F_0 continua, while discrimination peaks were present or absent, depending on the tone languages (Chan, Chuang, & Wang, 1975; Wang, 1976; Abramson, 1977, 1979; Connell, 2000; Francis, Ciocca, & Ng, 2003; Hallé, Chang, & Best, 2004; Xu et al., 2006; Wu & Lin, 2008; Peng, Zheng, Gong, Yang, Kong, & Wang, 2010). For the languages of Thai and Mambila in which there are three to four level lexical tones, native listeners identified these tones categorically, whereas no peaks were present in the discrimination functions (Abramson 1977, 1979; Connell, 2000). However, Mandarin Chinese listeners showed not only the sharp boundary between the level tone and the rising or falling tone in the identification function, but also exhibited a peak discrimination across the tone boundary, indicating a typical categorical tone perception (Chan et. al, 1975; Wang, 1976; Francis et al., 2003; Hallé et al., 2004; Xu et al., 2006; Wu & Lin, 2008). These findings imply linguistically-related auditory processing of F_0 contours. In addition, discrimination of contoured tones versus level tones is categorical, but discrimination of high level tones versus low level tones is not, indicating the effects of F_0 contours (e.g., level and rising/falling). Development of categorical perception in children learning tone languages should, accordingly, help them learn the location of tone category boundaries to master the many-to-one mapping between acoustic patterns and phonological categories. Cross-linguistically, adult listeners in tone languages perceived lexical tone in a more categorical manner than adult non-tone language listeners reflected by steeper identification slopes and greater peakedness in discrimination functions (Hallé et al., 2004; Xu et al., 2006; Wu & Lin, 2008; Peng et al., 2010).

Although categoricality of lexical tone perception has been examined for adult tone and non-tone language listeners (Halle et al., 2004; Xu et al., 2006; Wu & Lin, 2008; Peng et al., 2010), it is unknown whether lexical tones can be perceived categorically in school-aged children (6-8 years) who are still developing their perceptual strategy for tone perception at lexical and/or intonational levels. The goal of the present study was to examine how lexical tone was perceived for this population with different language backgrounds. In addition, previous studies indicated the importance of tone language exposure on tone contour perception (Peng et al., 2010), thus another goal of this study was to compare tone perception between Mandarin monolingual and English-Chinese bilingual children. We hypothesize that children regularly exposed to Mandarin may show a clear Mandarin tone perceptual boundary, whereas children who did not have Mandarin exposure may not. In addition, categorical tone perception was hypothesized to be affected by the extent of Mandarin exposure. Children with more Mandarin exposure may show sharper categorical perception of Mandarin tones.

2 Methods

2.1 Stimuli

Two sets of target stimuli were constructed based on a Mandarin Tone 1 word [ma1] "mother", recorded from a male native speaker of Mandarin (the second author) with normal speech and hearing function. The selection of [ma] was based on the familiarity of this syllable in both English and Mandarin Chinese and also the two phonemes ([m] and [a]) exist in both languages. The consonant and vowel are among the most common phonemes in the early speech inventories of both Mandarin and English children (e.g., Stoel-Gammon, 1985; Davis & MacNeilage, 1990; Zhu & Dodd, 2000). The average F_0 of the word [ma] over the word duration (362 ms) was 150 Hz, One set of stimuli was a rising-to-level continuum, constructed by manipulating the onset F_0 of the recorded word [ma] ("mother") from 100 to 150 Hz with the offset F_0 frequency fixed at 150 Hz (the left panel of Figure 1). The offset F_0 was manipulated from 150 to 100 Hz to form the falling-to-level continuum with the onset F_0 frequency fixed at 150 Hz (the right panel of Figure 1). Both continua included 11 stimuli that were equally spaced at 5 Hz of the onset or offset F_0. That is, for the rising-to-level continuum, the onset F_0 frequency was manipulated at 100, 105, 110, and up to 150 Hz, while for the falling-to-level continuum, the offset frequency was manipulated at 100, 105, 110, and up to 150 Hz. The speech syntheses were conducted in STRAIGHT (Kawahara Masuda-Kastuse, & Cheveigne, 1999), a speech synthesizer using a pitch-adaptive method to generate high-fidelity speech (Liu & Kewley-Port, 2004; Assmann & Katz, 2005). The procedure to synthesize the speech stimuli with F_0 continuum is described as follows: first, the F_0 contour of the recorded word [ma] with Tone 1 was obtained by the analysis of STRAIGHT; second, the F_0 contour was replaced by one of the F_0 continuum and reloaded into STRAIGHT for speech synthesis with no changes in other acoustic features such as intensity contour and spectrogram. Two sets of stimuli with F_0 continuum were generated.

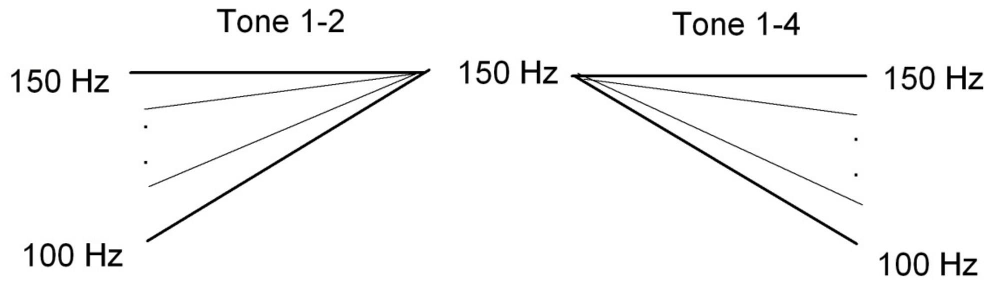

Figure 1. F_0 contours of Tone 1-2 (rising-to-level) continuum (left) and Tone 1-4 (falling-to-level) continuum (right) with a step size of 5 Hz.

2.2 Participants

Eight Mandarin Monolingual (MM), eight English Monolingual (EM) and eight English-Mandarin Bilingual (EMB) children participated in this study. MM children were recruited in the city of Handan in the province of Hebei at Mainland China that is located near Beijing. The average age of the MM participants was 7 years 2 months. All of the MM children spoke standard Mandarin as their native language at home and school and had no exposure to other Chinese dialects and any foreign language. EM children were recruited through a local school in Austin, Texas with an average age of 7 years. They spoke only English without exposure to any foreign language. EMB children were born in the US and recruited in Austin, Texas and in New York City through local Chinese schools. Their average age was 6 years 9 months. All EMB children started learning English after pre-school or kindergarten and Mandarin was the primary language (≥ 80%) used at home. The EMB children spoke both Chinese and English fluently with high intelligibility according to their parents' report and their informal conversations with the experimenters. Participants had no recorded history of speech, language, hearing or motor impairments according to parental report.

2.3 Procedures

Each participant completed two separate experimental tasks; 1.) identification and 2.) discrimination of tone stimuli. Each participant was tested separately in a quiet room at his/her home or at school[1], accompanied by one experimenter. Stimuli were presented by a Tucker Davis Technologies RM1 mobile sound processor to the left ears of the participants at 70 dB SPL through SONY MDR-7506 headphones[2]. The stimuli had 10-ms rise-fall ramps. The sound-pressure level measured in the NBS-9A 6-c[3] coupler by a Larson-Davis sound-level meter (Model 2800) with the linear weighting band set at 70 dB SPL for the speech signal.

A pseudo-identification paradigm (ABX) instead of categorization was used for the

---

[1] Noise level in the test site, home or school, was not measured. However, the signal-to-noise ratios in typical home and classroom environments range from +5 to +15 dB (Pearson, Bennett, and Fidall, 1977; Olsen, 1998).

[2] Due to the mono-output of the mobile sound processor (RM1), acoustic stimuli were delivered only to the left ears of listeners.

identification task, due to difficulty in training children, especially English monolinguals, to label Mandarin tones. The two extremes in the F_0 continuum of Mandarin tones served as the reference stimuli A and B. That is, for the rising-to-level continuum, the rising F_0 contour from 100 to 150 Hz (e.g., tone 2) and the flat $F_{0,}$ 150-150 Hz (e.g., tone 1) were the reference stimuli A and B. For the falling-to-level continuum, the falling F_0 contour from 150 to 100 Hz (e.g., tone 4) and the flat $F_{0,}$ 150-150 Hz (e.g., tone 1) severed as the reference stimuli A and B. The durations of the reference sounds (A and B) were the same as the target stimuli (X). The target stimuli from each of the continuums were presented in isolation with 10 repetitions in random order. The inter-stimulus interval was controlled at 500 milliseconds. Participants were asked to indicate which reference stimuli the target stimuli resembled. Thirty seconds were given for children to make a response after stimulus presentation. The next set of stimuli (ABX) was presented automatically in 1 second after a response was collected. The order of the stimulus presentation from the two sets of F_0 continuua was randomized for the identification experiment.

In order to make the experiments more interesting and accessible to young children, an interface panel with cartoon pictures (Mickey Mouse, Donald Duck, and Tweety Bird) was shown on a computer monitor. The pictures blinked when sounds were played. The experimenter instructed the participant as “Each of Mickey Mouse and Donald Duck will say a sound. Tweety Bird will imitate one of them. Please listen carefully, and choose the one that Tweety Bird imitates. Click the button below the picture using the computer mouse”.

The discrimination task used a three-interval, 2AFC (two alternative-forced-choice paradigm (ABA or AAB). A pair of sounds A and B separated by three steps (15 Hz) in each F_0 continuum (e.g., 100-115 Hz) was played following sound A. In half of the trials, ABA was presented while AAB was presented in the other half. There were eight pairs of stimulus comparisons (e.g., 100-115 Hz, 105-120 Hz, and up to 135-150 Hz) for each continuum. The inter-stimulus interval was set at 500 msec. The listeners’ task was to indicate which of the last two sounds (A or B) was different from the first sound (A). Each of the pairs was repeated 10 times and the presentation of these pairs was randomized. Altogether there were 80 trials (8 pairs x 10 trials per pair) for each continuum. Chidlren were provided thirty seconds to respond after stimulus presentation.. Similar to the identification tasks, cartoon pictures accompanied sound presentation in the discrimination task. The experimenter instructed the child as “Mickey Mouse will say a sound. Afterward, Each of Donald Duck and Tweety Bird will say a sound. Please listen carefully, and decide who speaks a different sound from Mickey Mouse; Donald Duck or Tweety Bird? Click the button below the picture”.

The identification task was completed first before the discrimination task for the same tone pair. A 15-minute training session was used before the tests to familiarize participants with the experimental procedure for both identification and discrimination tasks. Instructions were given in the participants’ native language for the monolinguals and in both English and Mandarin for the bilingual participants. Children were instructed to focus on the pitch of each stimulus for the identification and discrimination tasks. Short breaks were provided every 10 trials or as needed.

2.4 Data analysis

2.4.1 Identification functions

Identification scores of Tone 1 were measured at each onset (for Tone 1-2) and offset (for Tone 1-4) F_0 frequency and converted to percentages. For example, for Tone 1-2 identification, , when onset frequency of the stimuli is 140 Hz, if a participant identified 8 out of the 10 repetitions of the stimuli as Tone 1, the identification score at 140 Hz was 80%. Based on the binomial distribution of the identification scores and the sigmoid shape of the response function, a sigmoid model Eq. (1) between the identification score (y) and the onset/offset F_0 frequency (x) was fitted to obtain the identification function for each child. Goodness of fit was also evaluated to determine whether identification performance in each child fit the sigmoid model.

$$y = \frac{a}{1 + e^{-\frac{x - x_0}{b}}} + y_0 \quad (1)$$

where y_0 is the minimum identification score; a equals to maximum identification score minus minimum identification score; b and x_0 are the regression coefficients estimated by sigmoid model, representing the steepness factor and categorical boundary of the identification function, respectively.

The fitness of the sigmoid model was conducted in SigmaPlot® v10.0 that used the Marquardt-Levenberg algorithm (Marquardt, 1963) to find the parameters of the independent variable (x), giving the best fit between the equation and the data. An iterative process was conducted for the algorithm to seek the values of the parameters that minimized the sum of the squared differences between the values of the observed and predicted values of the dependent variable (y).

To compare the identification performance across groups, we obtained the parameters for individual participants whose identification performance yielded significant fitting for the sigmoid model. For each child in each F_0 continuum, categorical boundary x_0 was calculated from the value of onset/offset F_0 when identification score y equaled to 50%. The sharpness of categorical boundary was derived from the slope when identification score changes from 25% to 75% in the sigmoidal function.

2.4.2. Discrimination scores and related measures

Children's tone discrimination performance was measured by the discrimination score at each onset F_0 pair for the Tone 1-2 continuum and at each offset F_0 pair for the Tone 1-4 continuum. The discrimination score for a given frequency pair (onset or offset) referred to the percentage correct in the tonal discrimination task in ABA or AAB formats.

Following Xu et al. (2006), the discrimination scores for each child were then examined by three different measures: between category discrimination sensitivity (P_{bc}), within-category discrimination sensitivity (P_{wc}) and peakedness of the discrimination function (P_{pk}). P_{bc} was measured as the highest discrimination score near tone boundary obtained from the identification task. P_{wc} was calculated from the discrimination scores of the two comparison units at the ends of each tone continuum (e.g. 100 -115Hz and 135 -150 Hz). P_{pk} was estimated by the difference between P_{bc} and P_{wc}.

3 Results

3.1 Tone identification

Tone 1 identification scores for the Tone 1-2 continuum and Tone 1-4 continuum were averaged over the eight children within each of the three language groups. Figure 2 shows the average identification performance for the three groups of children in Tone 1-2 and Tone 1-4 continuum, respectively. For both Tone 1-2 and Tone 1-4 identification tasks, only the MM and EMB groups, displayed a sigmoid shape of response functions. Goodness of fit tests showed that all children in these two groups followed a sigmoid model in their identification performance (all $p < 0.05$). None of tonal identification for the EM children fit the sigmoid model (all $p > 0.05$).

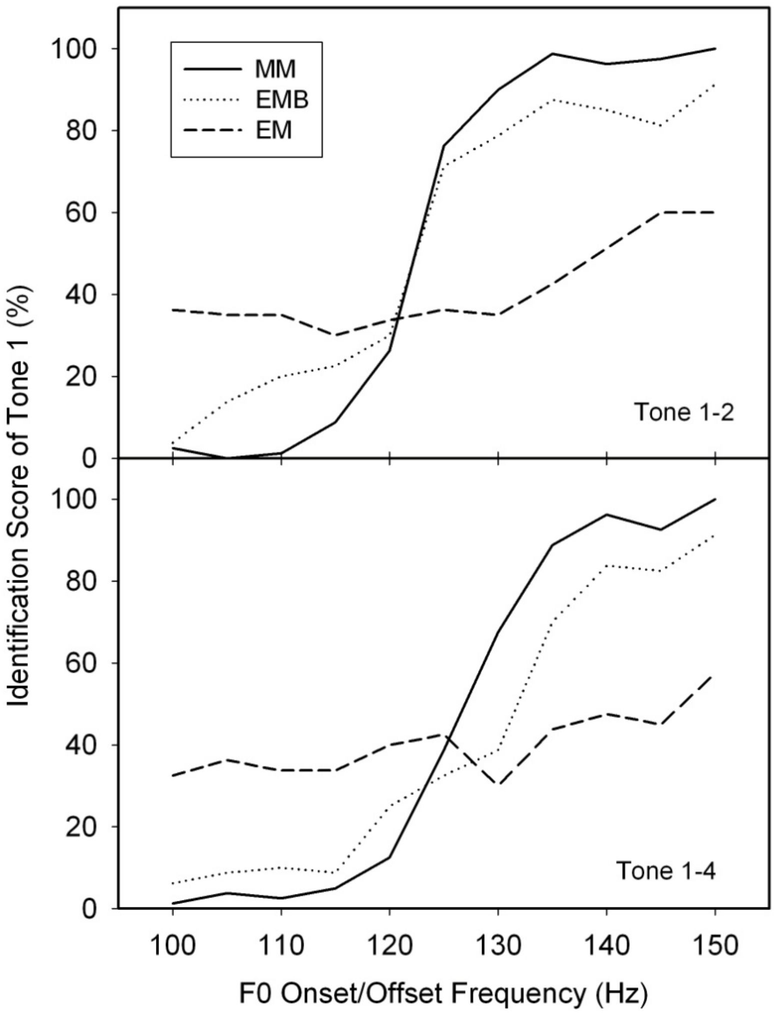

Figure 2. Average identification scores of tone 1 (percent correct) as a function of onset (Tone 1-2 continuum) and offset (Tone 1-4 continuum) F_0 frequency in Tone 1-2 continuum (upper panel) and Tone 1-4 continuum (lower panel) for MM, EMB, and EM children.

Categorical boundary and sharpness of tone identification were thus calculated only for children in the MM and EMB groups. Table I illustrates the average categorical boundary and slope (sharpness) for each group. Separate paired t-tests for Tone 1-2 and Tone 1-4 suggested that MM children displayed significantly sharper slopes than EMB children for both Tone 1-2 and Tone 1-4 identification (both $p < 0.05$). No significant difference was found for identification boundaries of Tone 1-2 and Tone 1-4 between the two groups (both $p > 0.05$).

Tone types	Child groups	Categorical boundary (Hz)	Slope (%/Hz)
Tone 1-2	MM	122.1	11.1
	EMB	121.5	6.7
Tone 1-4	MM	127.2	11.1
	EMB	130.1	7.5

Table 1. Average categorical boundary (at 50%) and sharpness (slopes) for MM and EMB groups.

3.2 Tone discrimination

Average discrimination scores of Tone 1-2 and Tone 1-4 over each child's individual scores within each of the three groups are shown in Figure 3. Interestingly, no prominent peaks were present across the tonal category for any of the three groups. As shown in Fig. 3, for the MM and EMB children, modest discrimination peaks were present near the tone boundary. For example, for the Tone 1-2 continuum, the discrimination peak was at the 140-125 Hz comparison for the MM children with the boundary of 122.1 Hz, while the peak was located at the 130-115 Hz comparison for the EMB children with the boundary of 121.5 Hz. For the Tone 1-4 continuum, the discrimination peak was present at the 140-125 Hz for the MM children with the 127-Hz boundary, while the EMB children showed the peaked discrimination at the 135-120 Hz comparison with the 130-Hz comparison. These results indicated that the location of modest discriminate peaks was matched with the tone boundary of the identification functions for the MM and EMB children.

As described above, three discrimination measures were derived from discrimination scores in reference to category boundaries obtained from identification tasks. Since none of the EM children showed category boundaries in the identification tasks, discrimination measures P_{bc}, P_{wc} and P_{pk} were only calculated from children in the MM and EMB groups. Table 2 shows the averaged discrimination measures within group. Paired t-tests did not show any significant differences between MM and EMB on any of the three discrimination measures (all $p > 0.05$).

Tone types	Child groups	P_{bc}	P_{wc}	P_{pk}
Tone 1-2	MM	77.5	58.1	19.4
	EMB	71.3	48.8	22.5
Tone 1-4	MM	78.8	54.5	21.8
	EMB	87.5	67.5	20

Table 2. Average tone discrimination measures (percent correct) for MM and EMB groups.

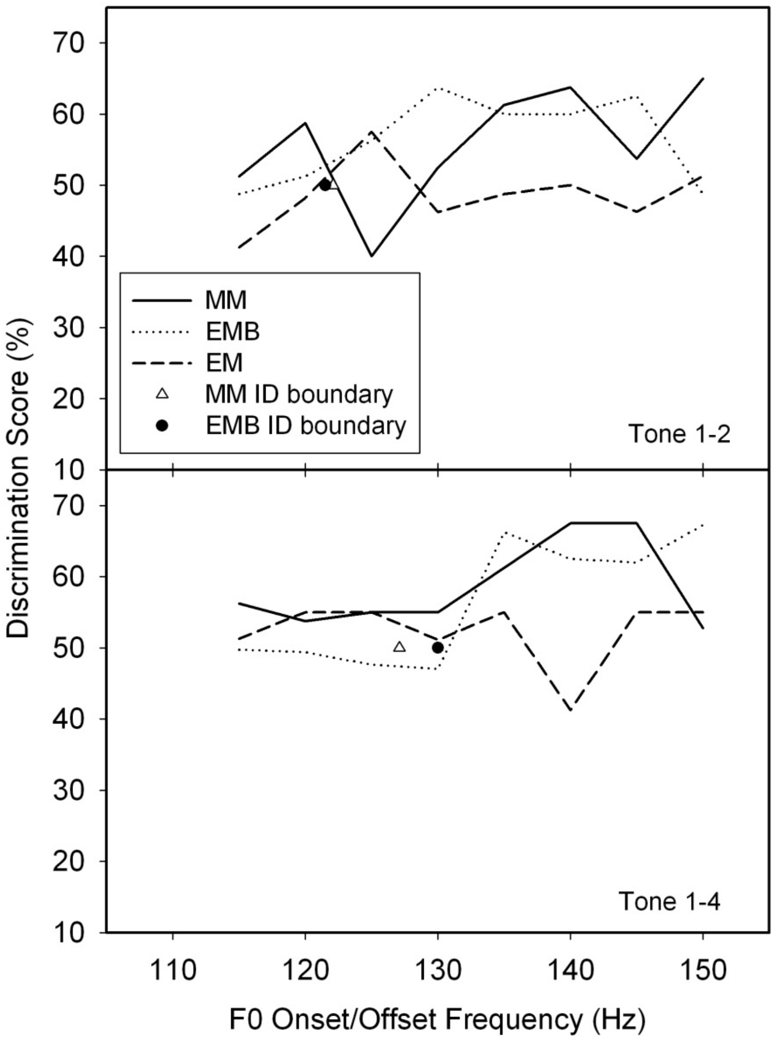

Figure 3. Average discrimination scores of tone discrimination in Tone 1-2 continuum (upper panel) and Tone 1-4 continuum (lower panel) as a function of the higher F0 frequency of each pair of tone contour comparisons for MM, EMB, and EM children. The identification boundary for the Tone 1-2 and Tone 1-4 continuums is represented for the MM (open triangle) and EMB children (closed circle).

4 Discussion

4.1 Tone identification

The perception of Mandarin Tone 1, 2 and 4 was investigated in three groups of young school-age children with varied language experience. Both MM and EMB groups exhibited a clear categorical boundary in tone identification tasks as expected. However, no categorical boundary was present in the identification responses from EM children. These results indicate that both groups of children with extensive Mandarin Chinese exposures were able to process lexical tones as linguistic categories, while the EM children were not. The lack of categorical boundary in children who have not been exposed to a tone language as opposed to results for non-tone exposed adults (Xu et al., 2006; Wu and Lin, 2008; Peng et al., 2010) might be because EM children were still in the process of developing auditory sensitivity towards tonal glides and pitch (Elliot et al., 1989; Jensen & Neff, 1993). Alternatively, EM children may develop tonal processing at larger prosodic levels before the syllabic level. The F_0 contour cue is developed in intonation production between 4 and 11 years old for English-learning children (Patel & Grigos, 2006), indicating that the EM children in the present study (6-8 years old) may not yet have fully developed the ability to differentiate intonation patterns using F_0 contour cues. Thus, at the lexical level, it is not surprising that these EM children were not able to use the F_0 cues efficiently to identify tone contours. Such capacity to perceive tone contours at the lexical level may be developed at elder ages for EM children, i.e., English-native adult listeners were able to identify the F_0 continuum categorically (Xu et al., 2006; Wu & Lin, 2008; Peng et al., 2010). Further research is needed to investigate the effect of age and language exposure on children's tone perception at the lexical and intonational level.

Although both MM and EMB groups showed clear categorical boundaries, the slope of the identification curve around the categorical boundary was shallower in EMB children. Since both groups of children had exposures to Mandarin, they have begun to establish the associations between acoustic F_0 patterns and linguistically-relevant tonal categories. Therefore, they exhibited similar locations for categorical boundaries. However, because EMB children spoke Mandarin only at home, and had comparatively less Mandarin exposure than MM children who spoke Mandarin all the time, their identification categories were defined less sharply than MM children. Our results suggest that amount of tone language input determines the existence of categorical boundaries for relevant ambient language contrasts. The amount and frequency of such input influences the sensitivity of tone perception. In addition, the exposure to English, a non-tone language might also affect EMB children's perception in Mandarin (von Hapsburg & Bahng, 2009). Peng et al. (2010) also reported that adult tone language users (Mandarin and Cantonese) identified Mandarin tones more abruptly than adult non-tone language users (German), implying that listeners' tone language experience appeared to be a primary factor to determine their lexical tone perception, even when listeners' tone language did not match with the language of target stimuli. Based on these results, it might be informative to evaluate tone perception performance in a continuum of bilinguals with varied amount of Mandarin exposure or measure tone perception across listeners of different tone and non-tone languages to quantify how tone language input influences sensitivity in tone identifications (Hallé et al., 2004; Peng et al., 2010).

4.2 Tone discrimination

Children in both MM and EMB groups showed typical boundaries in tone identification tasks. However, no prominent discrimination peaks were found near the observed identification boundary. As defined by Liberman et al. (1957), typical categorical perception should exhibit a sharp category boundary and a corresponding discrimination peak around boundary as predicted. In this study, categorical perception was present in a non-strict sense in children even with extensive linguistic experience in Mandarin, mainly due to non-typical tone discrimination functions. Even for adult listeners, the discrimination peaks across the tone boundary were present somewhat modestly (e.g., less than 85%) for Chinese listeners (Xu et al., 2006; Wu & Lin, 2008, Peng et al., 2010). In addition, the non-typical discrimination pattern in intonation perception (e.g., no or modest discrimination peaks across the boundary of intonation contrasts), which was primarily determined by F_0 contours, was also found for English and Chinese listeners (Ladd & Morton, 1997; Liu & Rodriguez, 2012). These finding suggested that tone pitch perception may be not categorically perceived, instead may be categorically interpreted at the lexical and intonation levels (Ladd & Morton, 1997).

Another possibility to modest discrimination peaks for the MM and EMB children might be due to the fact that the physical F_0 difference (15 Hz) used in the discrimination experiment was too small for these children to detect a difference. As suggested by a recent study (Liu, 2012) exploring the just noticeable difference (JND) of F_0 contour, Mandarin and English adult listeners required nearly 10 Hz to detect a tonal pitch change. Given that young children generally needed greater JND in frequency discrimination and tonal glide discrimination than adult listeners (Elliot et al., 1989; Jensen & Neff, 1993), school-aged children may need a frequency difference of more than 15 Hz used in this study to perform the task more reliably.

It should also be noted that the function of tone discrimination showed great listener variability (e.g., the discrimination peak was present at diverse F_0 values) for MM and EMB children in the present study. Such individual variability suggests that 6-8 years old children with extensive Mandarin Chinese experience may still develop their perceptual strategy, e.g., psychophysical and lexical processing, to discriminate lexical tone differences. The strategy to discriminate lexical tones with F_0 differences may be stabilized, i.e., adult-like, in older ages such as 8-to-11 years old. During this developmental period, the use of duration and F_0 cues in the production of lexical tones started to reach adult proficiency with reduced individual variability (Yang et al., 2008).

4.3 Application of this study

Results of this study indicated that categorical perception of Mandarin lexical tones was dependent on children's tone language experience. Children of 6-8 years old with plenty of Mandarin Chinese experience were able to perceive Mandarin tones categorically, while children without any tone language experience were not. As described above, EM children of 6-to-8 years old are still developing their strategy to use F_0 cues for intonation contrasts in their native language (English) and they may not pay attentions to F_0 cues at the syllabic level. Thus, if one's goal is to train EM children to learn Mandarin Chinese, training them to focus on F_0 cues at the syllable level will be critical.

5 Conclusions

The present study investigated categorical perception of Mandarin tones in children with different language backgrounds. In tonal identification, both MM and EMB children showed abrupt changes across tonal boundaries EM child listeners did not. In addition, MM children showed sharper tonal identification boundaries than EMB children. Only MM and EMB children showed moderately-peaked performance in tonal discrimination, probably due to small acoustic differences in pitch contour. Altogether, the MM and EMB children showed categorical tone perception while the EM children did not, although the categoricality was present in a non-strict format mainly because of the modest discrimination peaks across tone boundary.

Acknowledgements

Part of the study was presented at the 156th Acoustical Society of American Meeting. Collection of child data in China was made possible by the partial support of a Continental Airline Travel Grant awarded to the first author through College of Communication, the University of Texas at Austin. Special thanks are given to Barbara Davis for her comments on the earlier drafts of this manuscript. We would also like to thank Dr. Min Zhang and two reviewers for their constructive comments on the earlier drafts of this study. We gratefully acknowledge the children and their families for their time and cooperation.

6 References

Abramson, A. S. (1977). The noncategorical perception of tone categories in Thai. *Journal of Acoustic Society of America, 61*, S66 (A).

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

Assmann, P. and Katz, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *Journal of Acoustic Society of America, 117*, 886-895.

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tone. *Studies in Linguistics*, 22, 51-57.

Chao, Y. R. (1968). A Grammar of Spoken Chinese. *Berkeley: University of California Press*, pp. 1-28.

Connell, B. (2000). The perception of lexical tone in Mambila. *Language and Speech, 43*, 163-182.

Davis, B.L., & MacNeilage, P.F. (1990). The acquisition of vowels: a case study. *Journal of Speech and Hearing Research*, 33, 16-27.

Eimas, P., Sequeland, E.R., Jusczyk, P., and Vigorito, J. (1971) Speech perception in infants. *Science, 171*, 303-306.

Elliot, L. L., Hammer, M. A., School, M. E., and Wasowicz, J. M. (1989). Age differences in discrimination of simulated single-formant frequency transitions, *Perception and Psychophysics*, *46*(2), 181-186.

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

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). Identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171-189.

Gandour, J. (1978). The Perception of Tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41–76). New York: Academic Press.

Gandour, J., and Harshman, R. A. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1–33.

Gao, M. (2002). Tone in whispered Chinese: articulatory features and perceptual cues. *Master Thesis*. University of Victoria, Canada.

Hallé, P.A., Chang, Y-C., and Best, C.T. (2004). Identification and discrimination of Mandarin Chinese Tones by Mandarin Chinese v.s. French Listeners. *Journal of Phonetics, 32*, 395-421.

Huang, T. (2004). Language-specificity in auditory perception of Chinese tones. Unpublished doctoral dissertation. The Ohio State University, Ohio.

Jensen, J. K., and Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science, 4*, 104-107.

Kawahara, H., Masuda-Kastuse, I., and Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F_0 extraction: possible role of a repetitive structure in sounds. *Speech Commutation, 27*, 187-207.

Kent, R. D., and Vorperian, H. K. (1995). Anatomic development of the craniofacial-oral-laryngeal systems: A review. *Journal of Medical Speech-Language Pathology*, *3*, 145–190.

Kiriloff, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica, 20*, 63-67.

Ladd, D. R. and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical ? *Journal of Phonetics, 25*, 313-342.

Lehiste, I (1970). *Suprasegmentals*. Cambridge, MA: MIT Press, pp 1-202.

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.

Liu, C. (2012). Just-noticeable-difference of tonal pitch changes in speech and nonspeech stimuli for Chinese- and English-native speakers. *Journal of Acoustic Society of America,* submitted.

Liu, C., and Kewley-Port, D. (2004). STRAIGHT: a new speech synthesizer for vowel formant discrimination. *Acoustical Research Letters Online (ARLO), 5*, 31-36.

Liu, C., and Rodriguez, A. (2012). Categorical perception of intonation contrasts: effects of listeners' language background, *Journal of Acoustic Society of America, 131,* EL427-433.

Marquardt, D.W. (1963). An Algorithm for Least Squares Estimation of Parameters. *Journal of the Society of Industrial and Applied Mathematics*, *11*, 431-441.

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

Patel, R., and Grigos, M.I. (2006). Acoustic characterization of the question-statement contrast in 4, 7 and 11 year old children. *Speech Communication, 48*, 1308-1318.

Patel, R., and Brayton, J. (2009). Identifying Prosodic Contrasts in Utterances Produced by 4, 7, and 11 Year Old Children. *Journal of Speech Language and Hearing Research*, *52*(3), 790-802.

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., and Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38,* 616-624.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Cross language study of vowel perception. *Language and Speech*, *12*, 1-23.

Stoel-Gammon, C. (1985). Phonetic inventories, 15–24 months: a longitudinal study. *Journal of Speech and Hearing Research, 28*, 505–512.

von Hapsburg, D., and Bahng, J. (2009). Effects of noise on bilingual listeners' first language speech perception. *Perspectives on Hearing and Hearing Disorders: Research and Diagnostics , 13*, 21-26.

Wang, W. S.-Y. (1976). Language change. *Annals of the N.Y. Academy of Science, 208,* 61–72.

Wang, Y., Spence, M.M., Jongman, A., & Sereno, J.A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649-3658.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. J. *Child Language*, *31*, 749-778.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223-231.

Xu, Y., Gandour, J., and Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of Acoustic Society of America*, *20*, 1063-1074.

Wu, X., and Lin, H.(2008). Perception of Mandarin tones by Chinese and English Listeners. *Journal of Chinese Language and Computing*, *18*, 175-187.

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

Yuan, J. (2004). Perception of Mandarin intonation. *International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 25-48.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* , *27*, 3-42.