

The Effect of Acoustical Similarity on Lexical-Tone Perception of One-Year-Old Mandarin-Learning Infants

Feng-Ming Tsao

Department of Psychology, National Taiwan University

MS No.: 07031; Received: August 17, 2007; Revision: March 11, 2008; Accepted: April 17, 2008

Correspondence Author: Feng-Ming Tsao, Department of Psychology, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei, 106, Taiwan. (E-mail: tsaosph@mail2000.com.tw)

Speech perception abilities undergo rapid changes around the first birthday, such as developing the language-specific phonetic perception and word segmentation, but only a few studies have examined the development of lexical-tones, an essential phonetic unit to distinguish the meanings of syllables in Mandarin Chinese. This study explored the native tone perception of one-year-old Mandarin-learning infants. Specifically, this study addressed whether the acoustic similarity between lexical tones affected infants' perceptual discrimination performance. Infants ($n = 109$) were tested with the conditioned head-turn procedure when discriminating tone contrasts, Tone 1 vs. 3, 2 vs. 3, and 2 vs. 4, varying in the similarity of average fundamental frequency (F0) and F0 contour. Results showed that infants better discriminated tone contrast with greater acoustical differences (Tone 1 vs. 3), but they were less accurate discriminating acoustically similar tone contrasts (Tone 2 vs. 3 and Tone 2 vs. 4). Directional asymmetry was evident with the Tone 1 vs. 3 contrast, as the change from the background Tone 1 syllable to the target Tone 3 syllable was easier than the reverse. The results suggest that auditory processing is the essential mechanism for one-year-old infants to perceive native lexical tones, and that the perceptual mechanism for lexical tone in infants differs from Mandarin-speaking adults.

Keywords: *native lexical-tone perception, infant speech perception, acoustical similarity of tone, directional asymmetry*

Author Note

Feng-Ming Tsao, Department of Psychology, National Taiwan University, TAIWAN. This research was supported by a research grant from the National Science Council, Taiwan, to Feng-Ming Tsao. I thank Professor Patricia K. Kuhl (University of Washington, USA) for her technical support on setting up the infant test, Professors Huei-Mei Liu (National Taiwan Normal University) and Yue Wang (Simon Fraser University, Canada), for their helpful comments on preparing this manuscript, Yi-Ju Hsiao for testing infants, and all parents and infants participating in this study.

Correspondence concerning this article should be addressed to Feng-Ming Tsao, Department of Psychology, National Taiwan University, Taipei, Taiwan. E-mail: tsaosph@ntu.edu.tw.



Infants begin life with a language-general ability of speech perception, differentiating both the phonemes from their ambient language (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) and the phonemes from foreign languages until about 7 months of age (Kuhl et al., 2006; Tsao, Liu, & Kuhl, 2006; Werker & Tees, 1984). Adult listeners, on the other hand, are often perceptually confused with non-native phonemes (e.g., Trehub, 1973). Around 12 months of age, infants begin to demonstrate language-specific speech perception, i.e. the reduction of phonetic sensitivity to non-native phonetic contrasts (Best & McRoberts, 2003; Werker & Tees, 1984), while phonetic sensitivity to the native language improves (Kuhl et al., 2006; Tsao et al., 2006; Sebastián-Gallés, 2006).

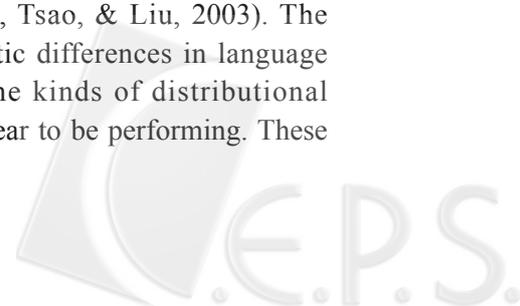
The onset age of language-specific speech perception is repeatedly reported to be around 12 months of age, using both behavioral (Kuhl et al., 2006; Tsao et al., 2006; Werker & Tees, 1984) and electrophysiological measures (Cheour, 1998; Minagawa-Kawai, Mori, Naoi, & Kojima, 2007; Rivera-Gaxiola, Klarman, Garcia-Sierra, & Kuhl, 2005). However, some studies failed to observe any phonetic sensitivity change for native and non-native languages around 12 months of age (Best & McRoberts, 2003; Polka, Colantonio, & Sundara, 2001), and the perceptual sensitivity for native language improves after the first year of life (Polka et al., 2001; Sundara, Polka, & Genesee, 2006). It is not clear why there is an age variation in the perceptual development for native and non-native languages. Without tests on additional phonetic contrasts, the issue whether 12 months is the onset age of language-specific processing cannot be answered.

Young infants distinguish the subtle acoustic difference between phonemes. 1- to 4-month-old English-learning infants were shown to discriminate a 20-ms difference in voice-onset-time (VOT), an important acoustic parameter to distinguish English voiced (/b, d, g/) from voiceless (/p, t, k/) stop consonants (Eimas et al., 1971). Subtle acoustic differences among vowel sounds are also discriminated, allowing infants to distinguish

the vowels of many languages early in life (e.g., Aldridge, Stillman, & Bower, 2001; Trehub, 1973). There is increasing evidence that in the first year of life, infants are acquiring detailed information about language by listening and analyzing linguistic input (Jusczyk, 1997; Kuhl, 2004; Werker & Tees, 1999). A variety of studies show, for example, that infants' exposure to ambient language results in rapid learning. By 6 months of age, infants engage in a detailed analysis of the distributional properties of sounds contained in the language they hear, and this alters their perception and produces more native-like phonetic processing (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Maye, Werker, & Gerken, 2002).

By 8-10 months of age, infants are capable of segmenting distinct lexical items from a continuous stream of speech by detecting transitional probabilities between syllables (Goodsitt, Morgan, & Kuhl, 1993; Saffran, Aslin, & Newport, 1996) and are able to detect the synchronization between the visual input periodicity and speech inputs (Hollich, Newman, & Jusczyk, 2005). At 8-9 months of age, infants are sensitive to the phonotactic and prosodic rules governing words, responding to the probability of occurrence of phonetic sequences (Johnson & Jusczyk, 2001; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Mattys, Jusczyk, Luce, & Morgan, 1999). Thus, these results suggest that between 6 and 9 months of age, with increasing linguistic experience, infants improve their perceptual processing of lexical units.

By 10-12 months of age, a developmental change of phoneme perception is apparent. Consonant discrimination shows an improvement for native phonemes, but a steep decline for non-native phonemes, reflecting a change that depends on linguistic experience (Kuhl et al., 2006; Rivera-Gaxiola et al., 2005; Tsao et al., 2006; Werker & Tees, 1984). A brief exposure to a non-native language alters this developmental decline for non-native phonemes (Kuhl, Tsao, & Liu, 2003). The ability to discern phonetic differences in language input is essential for the kinds of distributional analyses that infants appear to be performing. These



results show that listening experience with the native language would alter the perceptual abilities to detect phonetic segments and lexical units between 6 and 12 months of age.

Despite the extensive literature available on infants' perception of phonetic segments (e.g., vowels and consonants) and suprasegments (e.g., prosody), very few studies have examined the perceptual development of lexical tones in infancy. Harrison (2000) could be the first published study on infant lexical-tone perception, showing that 6- to 8-month-old Yorùbá (a tone language spoken in western Nigeria) infants discriminated Yoruba tone categories better than English-learning infants, but the issue of perceptual development for native tones in infancy was not addressed. The results of one recent study demonstrated that a perceptual decline for non-native tones occurred between 6 and 9 months of age when English-learning infants were tested with the lexical tones of Thai, but Mandarin-learning infants did not show any perceptual change for the non-native tone contrasts of Thai (Mattock & Burnham, 2006). However, it is still not clear whether Mandarin-learning infants demonstrate similar perceptual sensitivity for various lexical-tone contrasts of their native language around 12 months of age.

There are four tones in Mandarin, and phoneticians describe the perceptual representation of the Mandarin tonal contours as a high-level tone (Tone 1, ˥), a high-rising tone (Tone 2, ˨˨), a low-dipping tone (Tone 3, ˨˨˨), and a high-falling tone (Tone 4, ˥˨) (Chao, 1948; Howie, 1976). The contour shape and average fundamental frequency (F0) are two major acoustic correlates of lexical tones (Gandour & Harshman, 1978; Howie, 1976; Liu, Tsao, & Kuhl, 2007; Tseng, 1990; Wu, 1986). Figure 1 illustrates the F0 contours and averages of Mandarin lexical tones. The F0 contour of a syllable distinguishes its lexical meaning, but contour similarity between lexical tones varies. For example, Tones 2 and 3 exhibit similar F0 contours: both have a dynamic F0 shape. In contrast, Tones 1 and 3 exhibit dissimilar F0 contours: level vs. dynamic shape. The 'turning point,' the point in time at

which the pitch contour changes from falling to rising, is the acoustical measure for pitch contour, and is perceptually effective to differentiate Tone 2 from Tone 3 (Wang, Jongman, & Sereno, 2003).

In addition to the F0 contour, the average F0 (i.e., pitch height) is another acoustic dimension for Mandarin-speaking adults to perceptually organize lexical tones (Gandour & Harshman, 1978), and the order of average F0 (from high to low) in lexical tones is as follows: Tone 1 \approx Tone 4 > Tone 2 > Tone 3 (Liu et al., 2007). Thus, the average F0 difference between tones varied with tone contrast, for example, the Tone 1 vs. 3 contrast exhibits the largest average F0 difference, and the Tone 2 vs. 3 contrast exhibits a relatively small average F0 difference. With regards to both the average F0 and contour shape, the Tone 1 vs. Tone 3 contrast exhibits the greatest acoustic differences between lexical tones. On the contrary, the Tone 2 vs. Tone 3 contrast represents the most acoustically similar contrast. The average F0 differences between these tones are relatively small, and both tones exhibit the dynamic F0 contour (as illustrated in Figure 1). Other tone contrasts exhibit less acoustical similarity than the Tone 2 vs. 3 contrast. For example, Tones 2 and 4 have a relatively small average F0 difference but exhibit a large difference in their F0 contour turning points (Liu et al., 2007) and F0 contours: rising in Tone 2 and falling in Tone 4 (Figure 1). The acoustical similarity could be the reason that the Tone 2 vs. 3 contrast is frequently confused for non-tone language speakers (Chandrasekaran, Krishnan, & Grandour, 2007; Wang, Spence, Jongman, & Sereno, 1999) and even Mandarin listeners (Moore & Jongman, 1997; Shen, Lin, & Yan, 1993). For English adult speakers, the Tone 2 vs. Tone 4 contrast is less confusing than the Tone 2 vs. Tone 3 contrast, but this contrast is more difficult than the Tone 1 vs. Tone 3 contrast (Wang et al., 1999).

The acoustic similarity between tones may be associated with the slower acquisition of Tones 2 and 3 in Mandarin-learning children. Tones 2 and 3 were reported to be more frequently misarticulated than Tones 1 and 4 by Mandarin-learning children aged 1; 6 to 3; 0 (Li & Thompson, 1977). Moreover,

for 3-year-old Mandarin-speaking children, Tone 3 was still the most difficult tone to identify and produce (Wong, Schwartz, & Jenkins, 2005). Could the acoustic similarity between tones greatly affect infants' ability to distinguish tone contrasts? If this is the case, the general auditory ability that is not language-specific could be one of the major mechanisms for developing lexical-tone perception in Mandarin-learning infants. The perceptual sensitivities to tone contrasts in Mandarin-learning infants were predicted to vary with the acoustical similarity between tones, and the order of perceptual sensitivity would be Tone 1 vs. 3 > Tone 2 vs. 4 > Tone 2 vs. 3. In brief, the primary goal of this study was to examine the effect of acoustical similarity on the native lexical tone perception of Mandarin-learning infants around 12 months of age.

The second goal of this study was to examine the directional asymmetries in infant speech discrimination. Directional asymmetries are reported when infants are tested with a pair of syllables (e.g., syllables A and B) in a phonetic discrimination task, the accuracy of distinguishing phonetic change in one direction (e.g., target syllable: A, background syllable: B) is significantly better than in the other direction (e.g., target syllable: B, background syllable: A). For vowel discrimination, 6-8 and 10-12 month-old infants show directional asymmetries for both native and non-native vowel contrasts (Polka & Bohn, 2003). When American infants discriminate English /ra-/la/ contrast, the change from /ra/ to /la/ is more difficult for infants to detect than the reverse (Kuhl et al., 2006). The directional asymmetries could reflect either the general auditory perception or the linguistic bias (Polka & Bohn, 2003). Testing whether directional asymmetries exists in the discrimination of lexical tones in Mandarin-learning infants would assess the role of general auditory processing in development of lexical-tone perception.

In brief, this study aimed to examine the perceptual development of lexical tones in Mandarin-learning infants around 12 months of age. The effects of acoustical similarity and directional asymmetry on tone perception of one-year-old

Mandarin-learning infants were examined with three pairs of tone contrasts, varying in acoustical similarity, Tone 1 vs. Tone 3 (acoustically most distinct), Tone 2 vs. Tone 3 (acoustically most similar) and Tone 2 vs. Tone 4. The prediction was that the acoustical similarity between tone contrasts and the direction of presenting tone contrast could affect the perceptual discrimination performance and reveal the auditory basis of learning to perceive lexical tones in Mandarin-learning infants.

Method

Participants

The participants were 109 10-12 month-old Mandarin-learning infants, randomly assigned to three groups: Tone 1 vs. 3 group ($n = 33$, girls = 12, mean age = 11.43 months), among this group, target syllable = Tone 1 ($n = 15$, girls = 5), target syllable = Tone 3 ($n = 18$, girls = 7); Tone 2 vs. 3 group ($n = 36$, girls = 16, mean age = 11.02 months), among this group, target syllable = Tone 2 ($n = 18$, girls = 6), target syllable = Tone 3 ($n = 18$, girls = 10); and Tone 2 vs. 4 group ($n = 40$, girls = 21, mean age = 11.11 months), among this group, target syllable = Tone 2 ($n = 19$, girls = 9), target syllable = Tone 4 ($n = 21$, girls = 12). Pre-established criteria for inclusion in the study were that infants had no known visual or auditory deficits, were full term (born +/- 14 days from due date), had uncomplicated deliveries, had normal birth weights (2.72-4.5 kg.), were developing normally, and that members of their immediate families had no history of hearing loss.

Additional 27 infants failed to complete testing due to crying (3) or an inability to pass the training (24). Infants who failed to pass the training did not differentiate the phonetic contrasts: Tone 1 vs. 3 (drop-out rate = 8.3%), Tone 2 vs. 3 (drop-out rate = 28.0%), and Tone 2 vs. 4 (drop-out rate = 20.1%), $\chi^2(2, N = 136) = 5.09, p = .079$. However, the training passing rate differed by the direction of target syllable: Tone 1 \rightarrow Tone 3 (drop-out rate = 5.3%), Tone 3 \rightarrow Tone 1 (drop-out rate = 11.8%), Tone 2 \rightarrow Tone 3 (drop-out rate = 40.0%), Tone 3 \rightarrow Tone 2 (drop-out rate =

10.0%), Tone 2 → Tone 4 (drop-out rate = 22.2%), and Tone 4 → Tone 2 (drop-out rate = 17.4%), $\chi^2(5, N = 136) = 12.30, p = .031$, with more infants failing to pass the training in the direction of Tone 2 → Tone 3. Parents were paid NT\$400 for completing the experiment.

Mandarin-learning infants were recruited either through listings of names on the House Registry of the Da-An and Chung-Cheng Areas, Taipei City, Taiwan, or through the solicitation posted on the Internet. Although Taiwan is a multi-lingual society, Mandarin Chinese is the most dominant language in homes. The Mandarin-dominant (or only) language environments of the Taiwanese infants were verified through a language background questionnaire in Chinese that was administered to the caregiver before the study began.

Stimuli

The speech stimuli were /tɕ^{hi}1/ (duration = 690 ms), /tɕ^{hi}2/ (duration = 600 ms), /tɕ^{hi}3/ (duration = 770 ms) and /tɕ^{hi}4/ (duration = 482 ms) syllables, recorded by a female Mandarin-native speaker with a normal speaking rate in a sound-attenuation booth, and digitized with the speech analysis software, Computerized Speech Lab (CSL 4400) at a 22050 sampling rate, 16-bit resolution. Using naturally produced speech stimuli instead of the computer synthesized stimuli would provide the most natural tokens to examine the lexical-tone sensitivity in infants. As previously discussed in the introduction, these syllables were constructed

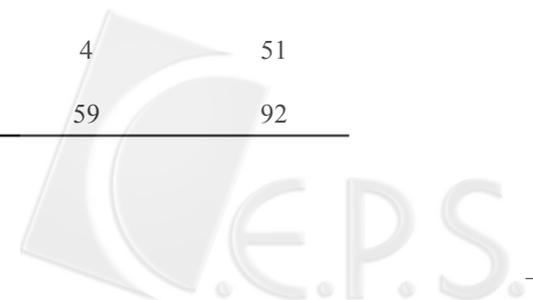
with three tone pairs varying in acoustic similarity with regards to the average F0 and F0 contour: (1) Tone 1 vs. 3 pair, the acoustically most distinct, (2) Tone 2 vs. 3, the acoustically most similar, and (3) Tone 2 vs. 4, the acoustical similarity in between. The duration, average F0, F0 range, and turning point [= (time of the minimal F0 ÷ tone duration) × 100%] are acoustical correlates of lexical tones (Liu et al., 2007). Table 1 lists the acoustical differences between lexical tones in each stimulus contrast and Figure 1 illustrates the average F0 and F0 contour of tone stimuli. Pitch measurements were taken from the vowel segments because the voiceless consonants of tone stimuli did not have vocal fold vibration and no pitch modulations would be produced during consonant segments. The duration of lexical tones is an acoustic correlate to distinguish tones in natural speech (e.g., Tseng, 1990). Thus, this acoustic parameter was preserved from the naturally-produced syllables in the digitized speech stimuli. The digitized speech samples were then edited with sound-editing software, Sound Forge 7.0, to equalize the RMS level of each syllable.

Apparatus

Speech stimuli were reproduced with 22050 Hz, 16-bit samples per second and presented by a computer (HP Compaq DC7100). The sounds were amplified (Yamaha RX V350) and delivered to infants in an adjoining sound treated test room via a loudspeaker (Bowers & Wilkins DM303). Parents and experimenters wore headphones (SONY

Table 1 Acoustical Differences between Lexical Tones in Each Tone Contrast for Testing Infants

Tone Contrasts	Acoustical Differences					
	Tone Duration (ms)	F0 Average (Hz)	F0 Range (Hz)	Turning Point (%)	F0 Initial (Hz)	F0 Final (Hz)
1 vs. 3	102	73	38	30	3	56
2 vs. 3	194	32	5	16.39	4	51
2 vs. 4	70	3	61	66.39	59	92



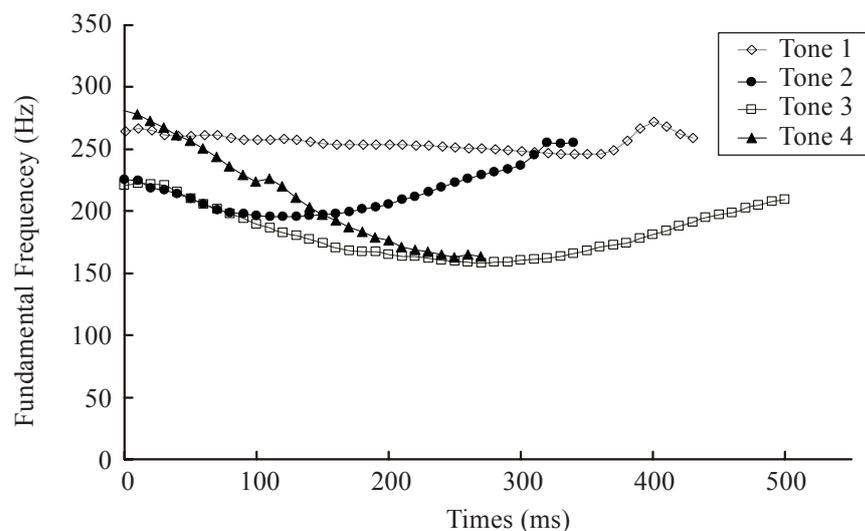


Figure 1. F0 contour of four lexical-tone stimuli (vowel segment) presented in this study.

MDR-CD 280) and listened to a music CD during the tests, so they could not distinguish between the stimuli presented to the infants. Infants' responses were monitored in the control room via the use of a digital camera (SONY Handycam PC350) and a video monitor. The computer used a data acquisition board (National Instrument PCI-6503) to activate the reinforcer and record infants' head-turn responses by an experimenter who pushed a button on a hand-held switch.

Test Suite

The test suite consisted of two rooms. In the sound-attenuation test room, an infant was held on its parent's lap, facing forward while an assistant sat at a 90-degree angle to the infant's right side. An assistant maintained the infant's attention by manipulating a series of engaging, silent toys to bring the child's gaze to midline (straight ahead of the infant). A bank of two visual reinforcers, located at a 90 degree angle to the infant's left side, each consisting of a dark plexiglass box (13" x 13" x 13") containing a commercially-available mechanical toy (e.g. a dancing snowman). The toys were not visible until activated, at which point lights mounted inside the box were illuminated. The visual reinforcers were placed on either side of the loudspeaker, and

were at eye level for the infant. A camera, located in front of the infant, fed an image of the test room to the adjoining control room, where an experimenter observed each infant's behavior.

Infant Testing Procedure

The Head-Turn (HT) technique was used to assess infants' discrimination abilities (Kuhl, 1985; Werker, Polka, & Pegg, 1997). Infants first were trained to produce a head turn for visual reinforcement whenever the "background" speech sound, e.g., /t^hi1/, repeated once every two seconds, would be changed to the "target" speech sound, e.g., /t^hi3/. The experimental protocol required a two-step training phase followed by a Test phase, all of which were under computer control.

The first step of the training phase consisted of Conditioning (+Intensity). During this phase, infants were trained to associate the presentation of the target speech sound with the activation of the visual reinforcers. The target sound interrupted the repetitive presentation of the background speech sound, and was presented at a level 4 dBA higher than the background speech sound. During the training phase, every trial was a target trial. The target stimulus was presented three times in a row. The onset-to-onset interstimulus interval was

2000 ms. The infant quickly learned to anticipate the visual reinforcer when the speech sound was changed from the background to the target. The infant had to respond to the sound change within 6 seconds after the first presentation of the target sound to watch the visual reinforcement. When the infant correctly anticipated the visual reinforcers with a head turn on two consecutive trials, the test proceeded to the next training phase, Conditioning (-Intensity).

In the Conditioning (-Intensity) phase, the target sound was presented at the same intensity level as the background sound; infants only used the phonetic difference between sounds as a cue. All other parameters of the experiment remained the same. Infants needed to correctly produce three anticipatory head turns to advance to the Test phase. Infants who failed to pass the two-phase training within 30 trials were eliminated from the experiment. The Test Phase consisted of 30 trials, an equal number of Change and Control (no-change) Trials, presented in random order. Infants were tested in 20-minute sessions on the same day.

In all phases of training and testing, trials were initiated by the assistant, who showed toys to the infant in the test room. The assistant initiated trials when infants appeared ready (focusing on the toys held by the assistant). The experimenter could not hear the stimuli presented during trials (a computer controlled gating network cut out the sound during a trial) and was unaware of the type of trial selected automatically by the computer. The experimenter judged the head-turn and pushed a button on a hand-held switch connected to the computer through the data acquisition board to indicate a head-turn. The assistant could not hear the stimuli being presented at any time during the experiment, but was informed that a trial was underway by a small light (out of the infant's view) that was automatically activated for the duration of a trial, necessary information for the assistant who was instructed not to change the toy in the midst of a trial.

Results

The head-turn response data for each infant during the test phase were summarized in terms of the four outcomes in a signal detection task: "Hits," "Misses," "False-positives," and "Correct rejections." Using this data, each infant's performance was converted to a percent correct, the measure of phonetic discrimination accuracy.

The first goal of this study was to examine whether the acoustical similarity between tone contrasts affected the native tone discrimination accuracy of one-year-old infants. The results of a two-way ANOVA (Tone contrasts \times Gender) provide the supporting data for the acoustical similarity effect on tone perception. The percent corrects of tone contrasts indicate that one-year-old infants perform significantly better discriminating the acoustically distinct Tone 1 vs. 3 contrast ($M = 73.39\%$, $SD = 12.73$) than Tone 2 vs. 3 contrast ($M = 60.74\%$, $SD = 12.67$) and Tone 2 vs. 4 contrast ($M = 57.81\%$, $SD = 9.58$), $F(2, 103) = 16.54$, $p < .001$, $\eta_p^2 = .24$. The LSD *post-hoc* test ($p < .001$) shows that the perceptual sensitivity order of tone contrasts: Tone 1 vs. 3 $>$ Tone 2 vs. 3 \approx Tone 2 vs. 4. There was neither a significant effect of Gender, $F(1, 103) = 1.04$, $p > .1$, nor Gender \times Tone interaction, $F(2, 114) = 1.82$, $p > .1$, showing the similar phonetic sensitivity for both boys and girls across tone contrasts. Table 2 lists the percent corrects of each tone contrast.

Another goal of the present study was to examine whether directional asymmetries exist when Mandarin-learning infants distinguish tone contrasts. The results of a one-way ANOVA on the percent corrects of six (= 2 directions \times 3 tone pairs) tone presentation conditions demonstrate the directional asymmetries, $F(5, 103) = 7.46$, $p < .001$, $\eta_p^2 = .27$. The results of LSD *post-hoc* test ($p < .05$) show the order of perceptual sensitivity of lexical tone contrasts: Tone 1 \rightarrow Tone 3 $>$ Tone 3 \rightarrow Tone 1 $>$ Tone 2 \rightarrow Tone 3 \approx Tone 3 \rightarrow Tone 2 \approx Tone 2 \rightarrow Tone 4 \approx Tone 4 \rightarrow Tone 2. Thus, the directional effect is shown only in one contrast, the Tone 1 vs. 3 contrast. Infants perform better detecting the tone

Table 2 Mean Percent Corrects (and Standard Deviations) of One-Year-Old Infants Distinguished Mandarin Lexical Tone Contrasts and Different Directions of Presenting Target and Background Syllables

Tone Contrast	Tone of Target Syllable	
	Tone 1	Tone 3
Tone 1 vs. Tone 3		
73.39% (12.74), <i>n</i> = 33	68.0% (10.53), <i>n</i> = 15	76.57% (13.37), <i>n</i> = 18
Tone 2 vs. Tone 3		
60.74% (12.67), <i>n</i> = 36	60.03% (12.78), <i>n</i> = 18	61.40% (12.89), <i>n</i> = 18
Tone 2 vs. Tone 4		
57.87% (9.58), <i>n</i> = 40	58.73% (7.28), <i>n</i> = 19	57.04% (11.40), <i>n</i> = 21

changes from the background /tɕ^hi1/ changing to the target /tɕ^hi3/ ($M = 76.57\%$, $SD = 13.38$) than the reverse ($M = 68.0\%$, $SD = 10.53$), LSD *post-hoc* test, $p = .036$. The directional asymmetry is statistically insignificant in other tone contrasts. The perceptual accuracy of Tone 1 vs. 3 is still higher than the perceptual accuracy of other tone contrasts, regardless of the tone presentation direction (LSD *post hoc* test, $p < .05$), showing that the Tone 1 vs. 3 contrast is perceptually more distinct than other tone contrasts. The percent corrects of each testing direction are listed in the Table 2.

Discussion

The goals of this study were to examine whether the acoustical similarity of tone contrasts could affect the native lexical-tone perception in one-year-old Mandarin-learning infants. The results show that the most acoustically distinct contrast (i.e., Tone 1 vs. 3) is the easiest contrast to discriminate, and the perceptual accuracy is similar for both the most acoustically similar contrast (i.e., Tone 2 vs. 3), and even the less acoustically similar contrast (i.e., Tone 2 vs. 4). Thus, the results demonstrate the effect of acoustical similarity on lexical-tone discrimination in Mandarin-learning infants. In addition, the results also show a directional asymmetry when infants are tested with the most acoustically distinct Tone 1 vs. 3 contrast, with Tone 1 → Tone 3 being the

easier direction for infants to discriminate. However, the perceptual sensitivity difference between the directions of tone presentation is not significant with other more acoustically similar tone contrasts (i.e., Tone 2 vs. 3 and Tone 2 vs. 4), showing no directional asymmetry of lexical-tone discrimination.

Despite many studies examining the developmental change of phonetic segment perception, only a few studies investigated the non-native lexical-tone perception of infants aged 6-9 months (Harrison, 2000; Mattock & Burnham, 2006). This may be the first study to examine infants' native tone perception at 12 months, an age when infants begin to demonstrate language-specific consonant perception (Kuhl et al., 2006; Tsao et al., 2006; Werker & Tees, 1984). The results of this study extend our understanding on the native lexical-tone perception development in infancy.

The result that the most acoustically distinct Tone 1 vs. 3 contrast is the easiest contrast than the other two contrasts (Tone 2 vs. 3 and Tone 2 vs. 4) for infants to discriminate demonstrates that the average F0 and contour differences between tones greatly affect the perceptual sensitivity of one-year-old Mandarin-learning infants. The results are consistent with previous findings that auditory difference limens are larger when infants are discriminating two tone sweeps from the same falling or rising direction than from either a tone sweep of the opposite direction (Aslin, 1989), and

also consistent with the prediction that a general auditory ability is essential to develop the lexical-tone perception in the first year of life. Recently, a neural network model on Mandarin lexical-tone acquisition shows that listening to the lexical tones produced by different speakers, without any feedback, is sufficient to develop the categorization of lexical tones (Gauthier, Shi, & Xu, 2007), suggesting the essential role of general auditory processing in learning tones.

However, the order of the easiness-of-tone discrimination in Mandarin-learning infants differs from English-learning adults who rely exclusively on their auditory ability to perceive lexical tones, suggesting that one-year-old Mandarin-learning infants utilize a different tone perceptual mechanism other than general auditory processing. For English speakers, a neurophysiological study showed that the peak latency of mismatched negativity (MMN) to Tone 1 vs. 3 contrast was shorter than Tone 2 vs. 3, demonstrating the easier perceptual distinction for Tone 1 vs. 3 (Chandrasekaran et al., 2007). In addition, English speakers discriminated Tone 2 vs. 4 contrast better than the Tone 2 vs. 3 contrast (e.g. Wang et al., 1999). However, Mandarin-learning infants appear to have a different perceptual sensitivity order than the English-speaking adults, with both Tone 2 vs. 3 and Tone 2 vs. 4 contrasts equally difficult for infants to discriminate.

What would account for the tone perceptual sensitivity order in Mandarin-learning infants? The average F0 and contours are essential to perceive lexical tones (Gandour & Harshman, 1978), and Mandarin-learning infants might show different perceptual weights for these acoustic cues. Although the results of this study cannot assess the relative contribution of acoustic cues in infant tone perception and no available study has examined this issue, the perceptual weights of English-speaking adults and the acoustical difference between lexical tones suggest that one-year-old Mandarin-learning infants weigh pitch height more than contour in the discrimination of tone contrasts. Using the multiple-dimensional scaling (MDS), the English speakers were reported to weigh the dimension of the average

F0 more than the contour, and Mandarin speakers showed the opposite when judging the similarity of lexical tone pairs (Gandour, 1983; Gandour & Harshman, 1978). The order of average F0 in Mandarin is Tone 1 \approx Tone 4 > Tone 2 > Tone 3 (Liu et al., 2007), showing a relatively larger average F0 difference in Tone 1 vs. 3 contrast. In addition, Table 1 shows that average F0 differences between tones of this study are larger for Tone 1 vs. 3 (73 Hz) than the other two tone contrasts, Tone 2 vs. 3 (32 Hz) and Tone 2 vs. 4 (3 Hz). With regards to the F0 contour of lexical tones, the F0 contours of Tones 2, 3, and 4 exhibit large F0 variations, but the F0 contour of Tone 1 is relatively flat (Liu et al., 2007). Table 1 also shows that the F0 range and contour turning point differences between tones are relatively larger in the Tone 2 vs. 4 contrast (F0 range difference = 61 Hz, turning point difference = 66.39%) than the Tone 2 vs. 3 contrast (F0 range difference = 5 Hz, turning point difference = 16.39%).

If the Mandarin-learning infants effectively perceive the F0 contour to discriminate the tone contrasts, the Tone 2 vs. 4 contrast should be easier than the Tone 2 vs. 3 because there are larger F0 range and contour turning point differences in the Tone 2 vs. 4 contrast than the Tone 2 vs. 3 contrast, and infants better discriminate tones with different pitch direction (e.g., Aslin, 1989). However, Mandarin-learning infants had a similar accuracy rate in discriminating the Tone 2 vs. 4 and Tone 2 vs. 3 contrasts, suggesting infants are not very effective to use the F0 contour cues to discriminate tones. As for pitch height, the average F0 differences between Tone 2 vs. 3 and Tone 2 vs. 4 are smaller than Tone 1 vs. 3 (Liu et al., 2007), and the average F0 differences between tones in the tone contrasts of this study (Table 1) show the same order of tone difference. If infants rely more on the average F0 than the contour difference to discriminate tones, infants would perform with better perceptual accuracy of the Tone 1 vs. 3 contrast than both the Tone 2 vs. 3 and Tone 2 vs. 4 contrasts. This is the order of infant tone sensitivity shown in this study.

Both average F0 and contours are changed across tones in naturally produced speech. This issue

of perceptual weight in Mandarin-learning infants could not be adequately addressed in the current study. A future study using computer-generated tone stimuli varying only in either average F0 or contour, and testing infants with the Tone 1 vs. 4 contrast, where both tones exhibit statistically similar average F0 but very large F0 range and turning point differences (Liu et al., 2007), could be the next step to examine this issue. The results of this study suggests that Mandarin-learning infants around 12 months of age utilized the similar tone perception processing strategies as non-tone language adults to perceive tones (Gandour, 1983; Gandour & Harshman, 1978), and their ability to perceive lexical tones is still developing after their first year of life.

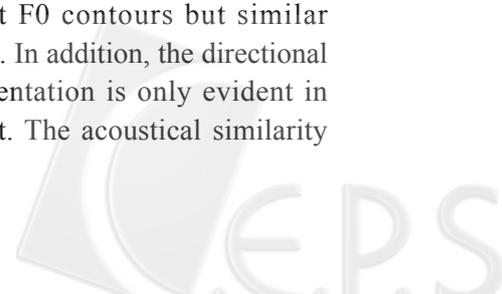
In addition to the acoustical similarity, the result that Tone 1 \rightarrow Tone 3 is the easier direction for infants to discriminate than Tone 3 \rightarrow Tone 1 demonstrates the directional asymmetry of lexical tone perception. The directional asymmetries of phonetic discrimination have been shown using vowels (Polka & Bohn, 2003) and semi-vowels (Kuhl et al., 2006) when infants are tested with the same testing procedure in the present study, but some studies failed to report the directional asymmetry. For example, 6-7 month-old English-learning infants did not show this asymmetry discriminating native vowels presented with both high pitch and low pitch (Trainor & Desjardins, 2002). Although there are inconsistent findings of directional asymmetry, what would account for the directional asymmetry shown in this study? General auditory perception could be one explanation. For example, the directional asymmetry is evident in adults discriminating tone glides, and the down-glide is more difficult to discriminate than the up-glides (Madden & Fire, 1997). Specifically, the auditory forward-masking effect could be one reason for the directional effect. The directional effect is only evident in the discrimination of Tone 1 vs. 3 contrast and not shown in other contrasts. The F0 contour patterns of Tone 1 (level) and 3 (dipping) are very different. The difficult direction of the tone contrast is /t ϵ ^{hi}3/ \rightarrow /t ϵ ^{hi}1/, suggesting that infants find it more difficult to detect the syllables with a relatively flat F0 contour

from varied F0 contours than the reverse, thus, the forward masking effect of the background Tone 3 on target Tone 1 would greatly reduce the possibility of infants detecting tone changes. On the contrary, when the Tone 1 is the background syllable, the forward-masking effect of Tone 1 on Tone 3 would not be too robust so as to reduce the possibility for infants to detect the sound change of the following Tone 3 syllables in the head-turning testing procedure. However, only the Tone 1 vs. 3 contrast is tested in this study, and more studies using the other level vs. contour tone contrasts, i.e., Tone 1 vs. 2 and Tone 1 vs. 4, would provide sufficient data to test whether the masking-effect difference between level and contour tones accounts for the directional asymmetry in tone perception.

Although the directional effect is only evident with the Tone 1 vs. 3 contrast, it would be interesting to learn whether there is a developmental trend for directional asymmetries in the lexical-tone perception. If the directional effect is not evident with younger infants, it would reveal that the listening experience with lexical tones influences the directional asymmetry of phonetic discrimination in older infants. In brief, the present study demonstrates the directional asymmetry in the discrimination of lexical tones and suggests that the auditory mechanism is essential to develop tone perception in one-year-old infants.

Conclusion

The one-year-old Mandarin-learning infants were tested with lexical-tone contrasts, varying in acoustical similarity and the direction of sound presentation. Results show that the lexical-tone discrimination performance of infants changes with the acoustic similarity, that is, better discrimination for a more acoustically distinct contrast (Tone 1 vs. 3) and less accurate discrimination of a more acoustically similar contrast (Tone 2 vs. 3) and tone contrasts with different F0 contours but similar average F0 (Tone 2 vs. 4). In addition, the directional asymmetry of tone presentation is only evident in the Tone 1 vs. 3 contrast. The acoustical similarity



effect and directional asymmetry shown in this study suggest that general auditory perception is essential for one-year-old Mandarin-learning infants to perceive native tones. In addition, Mandarin-learning infants may not utilize the similar perceptual weight of acoustic tone cues as Mandarin-speaking adults. The on-going infant studies are testing tone perception in younger infants (7-month-olds) and in English-learning infants to examine whether listening experiences with lexical tones could alter the perceptual organization of lexical tones.

References

- Aldridge M. A., Stillman R. D., & Bower T. G. R. (2001). Newborn categorization of vowel-like sounds. *Developmental Science*, 4, 220-232.
- Aslin, R. N. (1989). Discrimination of frequency transitions by human infants. *Journal of the Acoustical Society of America*, 86, 582-590.
- Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183-216.
- Chandrasekaran, B., Krishnan, A., & Grandour, J. T. (2007). Mismatch negativity of pitch contours in infants influenced by language experience. *Brain Research*, 1128, 148-156.
- Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.
- Cheour, M. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351-353.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149-175.
- Gandour, J. T., & Harshman, R. A. (1978). Cross-language differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21, 1-33.
- Gauthier, B., Shi, R., & Xu, Y. (2007). Simulating the acquisition of lexical tones from continuous dynamic input. *Journal of the Acoustical Society of America*, 121, EL190-195.
- Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, 20, 229-252.
- Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, 110, 581-616.
- Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, 76(3), 598-613.
- Howie, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548-567.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 252-293.
- Kuhl, P. K. (1985). Categorization of speech by infants. In J. Mehler & R. Fox. (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion* (pp. 231-262). Hillsdale, NJ: Erlbaum.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13-F21.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, USA*, 100, 9096-9101.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months

- of age. *Science*, 255, 606-608.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin – speaking children. *Journal of Child Language*, 4, 185-199.
- Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43, 912-917.
- Madden, J. P., & Fire, K. M. (1997). Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency. *Journal of the Acoustical Society of America*, 102, 2920-2924.
- Mattock, K., & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, 10, 241-265.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465-494.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-111.
- Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience*, 27, 315-321.
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864-1877.
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221-231.
- Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/ - /ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, 109, 2190-2201.
- Rivera-Gaxiola, M. C., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport*, 16(5), 495-498.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Sebastián-Gallés, N. (2006). Native-language sensitivities: Evolution in the first year of life. *Trends in Cognitive Science*, 10, 239-241.
- Shen, X., Lin, M., & Yan, J. (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *Journal of the Acoustical Society of America*, 93, 2241-2243.
- Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d - ð/ in monolingual and bilingual acquisition of English. *Cognition*, 100, 369-388.
- Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, 9, 335-340.
- Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9(1), 91-96.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and nonnative affricate-fricative contrasts: Cross-Language tests on adults and infants. *Journal of the Acoustical Society of America*, 120(4), 2285-2294.
- Tseng, C. Y. (1990). *An acoustic phonetic study on tones in Mandarin Chinese*. Taipei, Taiwan: Institute of History & Philology Academia Science.
- 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.
- Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033-1043.
- 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.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.
- Werker, J. F., Polka, L., & Pegg, J. E. (1997). The conditioned head turn procedure as a method

- for testing infant speech perception. *Early Development and Parenting*, 6, 171-178.
- 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, 1065-1079.
- Wu, Z. J. (1986). *The Spectrographic Album of Monosyllables of Standard Chinese*. Beijing, China: Social Science Press.



聲調相似度對漢語周歲嬰兒聲調知覺的影響

曹峰銘

國立台灣大學心理學系暨研究所

在周歲左右，嬰兒的語音知覺能力經歷快速的變化。嬰兒開始呈現語言特定的知覺處理歷程，例如，開始降低對外語的語音音段（例如，元音與輔音）敏感度，同時提高對母語語音及超音段（例如，音韻規律）的知覺敏感度。然而，到目前為止，只有少數的研究探討「聲調」，這個在聲調語言當中用來區分音節語義的重要語音單位。而且，尚無研究探索漢語為母語嬰兒的聲調知覺發展。本研究的目的在於探究嬰兒對母語聲調知覺的發展，特定而言，檢視聲調之間聲學特性相似的程度，是否會影響嬰兒對不同聲調對比區辨的敏感度。周歲大的嬰兒（ $n = 109$ ），參加以「轉頭步驟（head-turn procedure）」進行的聲調區辨實驗。其中一組分辨聲學特性差異最大的「一聲和三聲（ $n = 33$ ）」，另一組分辨聲學特性最接近的「二聲和三聲（ $n = 36$ ）」，還有一組區分聲學特性相似性在兩組對比之間的「二聲和四聲（ $n = 40$ ）」。結果顯示，一聲和三聲對比的區辨正確率（ $M = 73.39\%$ ），明顯地高於二聲和三聲（ $M = 60.74\%$ ）及二聲和四聲（ $M = 57.87\%$ ）的區辨正確率。此外，結果也顯示當嬰兒區辨一聲及三聲時，目標音節為「三聲」會比「一聲」的區辨正確率高，也就是呈現「方向不對稱性」。但是此一現象只出現於聲學特性差異最大的一及三聲對比，其他聲調對比沒有此一現象。研究結果呈現周歲嬰兒的漢語聲調知覺發展情形。而聲調的聲學相似性對區辨敏感度的作用，也意涵一般聽覺發展機制在嬰兒發展母語聲調知覺上的重要性，以及周歲嬰兒對聲調聲學線索的處理特性與說漢語的大人不同。

關鍵詞：嬰兒語音知覺、聲調發展、聲學相似性、方向不對稱性

