# Categorization of Lexical Tones in Mandarin-Learning Infants

*Jun Gao [1], Rushen Shi [2], Aijun Li [3]*

[1] Graduate School, Chinese Academy of Social Sciences, China
[2] Université du Québec à Montréal, Canada
[3] Institute of Linguistics, Chinese Academy of Social Sciences, China

`gaojungscass@126.com, shi.rushen@uqam.ca, liaj@cass.org.cn`

## Abstract

It is well-known that children are born with the perceptual ability to discriminate not only native phonetic contrasts, but also non-native ones. However, recent research suggests that not all contrasts are equally discriminable. Children cannot discriminate some acoustically similar phonetic contrasts early in infancy. They need a certain period of exposure with these contrasts to learn to discriminate them. In this study we inquired if similar lexical tones are discriminable in young infants. Mandarin-learning 4-13-month-olds' categorization of Mandarin rising and dipping tones were assessed in a visual habituation procedure. The results show that infants successfully categorized the two similar tones.

**Index Terms**: infant speech perception, language acquisition, categorization, lexical tones

## 1. Introduction

Mandarin has four lexical tones, high level (55, ˥), high rising (35, ˧˥, hereinafter referred to as "rising"), low dipping (214, ˨˩˦, hereinafter referred to as "dipping") and high falling (51, ˥˩) [2] (see Figure 1). Acoustically, rising and dipping tones are similar in contour. Both tones start at around the mid region of F0 range and end at the upper region of F0 range. The changing direction of F0 for both tones is from low to high, even though the dipping tone begins with a fall and then turns to rise. Furthermore, native speakers do not produce typical tokens of tones all the time. In atypical tokens of rising and dipping tones, the contrasts in ΔF0 (the F0 change from the onset to the turning point) and the timing of the turning point, which are cues that distinguish the two tones, may be weakened [17].
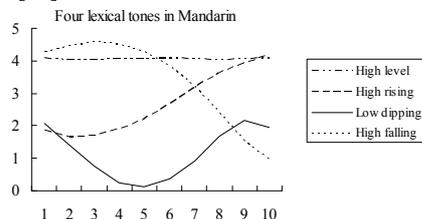


Figure 1: *Time-normalized F0 contours of the four lexical tones in Mandarin produced by a female native speaker.* (*data from Lee Sung Hoon, Graduate School, Chinese Academy of Social Sciences*)

The factor of acoustic similarity may give rise to the difficulty in the perception and production of rising and dipping tones. Native Mandarin speakers, who are usually perfect in perceiving all four tones, can nevertheless make mistakes when identifying rising and dipping tones [17]. The perception by listeners whose native language is not Mandarin, such as Hong Kong Cantonese and Japanese, exhibits confusions between rising and dipping tones, in addition to confusions between high level and rising, and between high level and high falling [19]. The production of rising and dipping tones, regardless of non-native learners' language background, is characteristic of errors in both register and contour, and the produced tones are undistinguishable to the native Mandarin ears [4, 11, 23].

The similar characteristics of rising and dipping tones also lead to their late acquisition by Mandarin-learning children [9, 10, 24, 25]. According to a longitudinal 7-month study by Li and Thompson [9] on the production data of 17 Mandarin-learning children from 1;6 to 3;0, "the high [level] and falling tones are acquired earlier and more easily than the rising and dipping tones" (p.189) and "c[C]onfusion persists in the form of substitution errors throughout Stages II and III between the rising and dipping tones" (p.189). In Liu's analysis of the production data of 10 18-23-months-old Mandarin-speaking children [10], incorrect substitutions of one tone for another were common before 2 years of age, with the rising tone tending to be the last one acquired by children. The chronological observation of Yang [25] on two children from 12 months to 25 months of age showed that the acquisition order of the four lexical tones was high level first, followed by falling and rising, and lastly dipping tone. The investigation by Wong et al. [24] on Mandarin-speaking children within the age range from 2;10 to 3;4 revealed that, of the four lexical tones, dipping tone was the one that exhibited most confusions in production.

Children's difficulty with the production of rising and dipping tones may be due to one of the following factors. One possibility is that children may be too young to master the control of articulators to produce the two tones, especially the dipping one, a complex tone. That is, children may perceive these two tones well but fail to produce them due to lack of appropriate motor control. Alternatively, the difficulty may be rooted in their poor perception of these two tones. Children may have difficulty discriminating the two tones because of their acoustic similarity. Of these factors, we are most concerned with the question of acoustic similarity – whether children have the perceptual ability to distinguish contrasts of similar tones. In light of the existing literature on phonetic perception during infancy, we inquire whether infants have the ability to discriminate all phonetic contrasts in place at birth, both acoustically similar and dissimilar, and later become attuned only to native contrasts, or whether infants can initially distinguish only highly contrastive phonetic categories while gradually developing the ability to perceive acoustically similar phonetic contrasts following accumulated exposure.

This question was addressed in a recent study on the perceptual development of segments [15]. Narayan et al. [15] showed that by 10-12 months of age, Filipino-learning infants

were able to discriminate the native contrast [n]-[ŋ], which is acoustically similar, while younger infants at 6-8 months of age failed to show the ability. These results suggest that similar phonetic categories are gradually established by continuous exposure to the contrasts in the ambient language. In another study on consonants, Kuhl et al. [6] showed that English-learning infants gradually improved in their discrimination of English /r/ versus /l/ (which are arguably acoustically similar) during the later half of the first year of life. Moreover, for some acoustically highly similar contrasts such as the coronal [d]-[ð], the development of discrimination needs to take a much longer period of learning, well into later childhood [16, 20]. The idea of acoustic similarity as a factor influencing the development of phonetic categories is supported by perception of acoustically highly dissimilar non-native contrasts in both children and adults. Notably, adults who are supposed to be insensitive to non-native contrasts can even discriminate non-native contrasts, such as Zulu clicks [1], which are highly dissimilar.

Findings from the lexical tone acquisition literature are consistent with the view of acoustic similarity. Lee et al. [7] studied Cantonese-speaking children aged from 2;09 to 3;03, showing that in early stages of learning, dissimilar tones were better discriminated than similar tones – high level (T1, see Figure 2) vs. low falling (T4) and high level (T1) vs. high rising (T2), which are dissimilar in F0 at the onset, were more distinguishable for children than high rising (T2) vs. low falling (T4), the onset of which is similar in F0. Lei [8] showed that Cantonese-learning 6-8-month-olds discriminated high level (T1) and mid-low level (T6), which are dissimilar in terms of the large difference in F0 height, and that infants had difficulty in distinguishing high level (T1) vs. mid level (T3), and mid level (T3) vs. mid-low level (T6), which are less different in F0 height. These two studies [7, 8] show that acoustically more similar tones are more difficult and are thus acquired later than dissimilar tones.
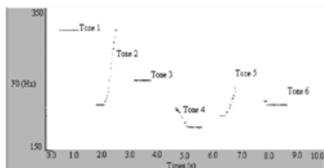


Figure 2: *F0 contours of six lexical tones in Cantonese. (cf. [7])*

With regards to tonal perception in Mandarin, Tsao and colleagues [22] tested Mandarin-learning infants of different ages (6-8 month vs. 10-12 month), to determine whether the development of tonal discrimination is related to the extent of acoustic similarity. A significant improvement was observed in the discrimination of the dissimilar pair, high level vs. dipping, by older infants (10-12 month olds) in comparison to 6-8-month-olds. The performance for the rising vs. high falling pair was poorer for both 6-8-month-olds and 10-12 month olds. In another study, Tsao [21] tested Mandarin-learning 10-12-month-old infants' discrimination of three tonal contrasts, and found that the rising vs. dipping and the rising vs. high falling contrasts both yielded poorer performance than did the high level vs. dipping contrast. Tsao and colleagues suggested that the poorer discrimination was due to the fact that both tonal pairs (rising vs. dipping; rising vs. high falling) were acoustically similar. For example, according to Tsao, rise and high falling tones are similar in terms of average F0 and may be more confusing. However,

given the experimental results from other studies [3, 5, 12, 13, 18], it is not clear whether these two tonal contrasts are indeed difficult for infants. Mattock and Burnham showed that 6- and 9-month-old Mandarin-learning and Cantonese-learning infants can discriminate Thai rising versus low tones and rising versus falling tones [12] (see Figure 3). In fact, Mattock and colleagues later tested one of these tonal contrasts (rising versus low) with infants aged 4-6 months from non-tonal language environment (English, French), and found that even these infants perceived this Thai tonal contrast [13]. In Shi [18] 8-11-month-old Mandarin-learning infants perceived the Mandarin rising and high falling tonal contrast even in highly variable tonal contexts. More strikingly, 2- to 3-month-old English-learning infants have been shown to discriminate single syllables bearing rise and fall intonations [5].
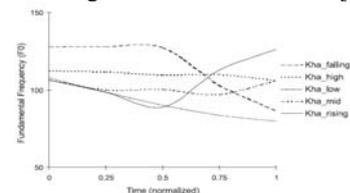


Figure 3: *Time-normalized F0 contours of the five Thai tones (spoken by a male Thai speaker). (cf. [12])*

These results suggest that the questions of similar and dissimilar tones as well as infants' initial tonal perception ability remain open. More experiments are therefore necessary to fully understand whether infants have the perceptual ability to discriminate similar tones and what constitute similar versus dissimilar tones.

The purpose of our research is to examine Mandarin-learning infants' perception of various tonal pairs in Mandarin. In this paper we report the results of one experiment on infants' categorical discrimination of acoustically similar rising versus dipping tones. We used a visual habituation paradigm, a procedure that was different from that used by Tsao and colleagues [21], who had previously tested this tonal contrast.

## 2. Method

### 2.1. Participants

20 monolingual Mandarin-learning infants aged 4 to 13 months participated in the experiment, 13 girls and 7 boys. Seven additional infants (2 girls and 5 boys, aged 7-12 months) were tested but were not included in the analyses due to fussiness and being out of the camera view (n=4), too-fast-switches of eyeballs (n=1) and experimenter errors (n=2).

### 2.2. Recording and stimuli

The tone-bearing syllable chosen was /tsʰan/, which agrees with the phonotactic rules of Mandarin phonology. The target syllable used for the experiment was [tsʰan] rising tone and [tsʰan] dipping tone, respectively. These are unfamiliar lexical items to infants of this age. A female speaker of Mandarin Chinese produced the stimuli in the infant-directed speech style in a sound-proof acoustic chamber. She produced multiple repetitions of the two target words, [tsʰan]-rising and [tsʰan]-dipping, while also producing repetitions of the syllable in other tones ([tsʰan]-high level and [tsʰan]-high falling) as well as producing other filler words with a different

syllable construct. The speech was recorded with a 22050Hz sampling frequency, 16 bit resolution.

For each target syllable, 13 tokens were selected based on the results of an identification test by 6 randomly-chosen Mandarin-speaking adults and on the distribution of duration. The tokens of the two target syllables were carefully selected such that they had the same duration distribution. Note that duration is one of the cues distinguishing rising and dipping tones in Mandarin. Thus, this control aimed at reducing the acoustic dissimilarity between the two tones. The mean duration of the rising tone tokens was 718ms (max = 806ms, min = 631ms) with the standard deviation of 63ms. The mean duration of the dipping tone tokens was 717ms (max = 802ms and min = 630ms) with the standard deviation of 63ms. All tokens were adjusted through Cool Edit Pro 2.0 into comparable amplitude. The visual stimulus was a colorful checkerboard-like geometrical image.

## 2.3. Apparatus

The experiment was conducted in a dimly-lit sound-proof chamber (3.25m*3.5m*2.4m), where the mother and the infant sat in a sofa facing an LG LCD display screen (M4212CF) about 1.7m in front of them. On each side of the LCD screen, there was a Genelec 1029A loudspeaker. The display screen and the loudspeakers were connected to a computer outside the chamber. Under the screen, a Panasonic low-light digital video camera (AG-HMC153MC) filmed the infant and transmitted the video to a screen monitor (LG W1942T) outside the chamber for online coding. The experimenter outside the testing chamber was blind to the stimuli of the experiment, and coded online the infant's looking to and away from the screen. The experiment was driven by a computer program, which presented the audio-visual stimuli contingent upon the infant's looking to the screen. Looking time data were recorded automatically. The software controlled the initiation and termination of all trials as well as the delivery of the audio-visual stimuli. During the experiment, the mother held the infant on her lap and was listening to music through headphones (Peltor HTM79A), which served to mask her from the audio stimuli presented to the infant. The mother was required not to disturb or interrupt the infant.

## 2.4. Procedures and design

Each infant was tested individually. Once the mother and the infant settled in the sofa, the experimenter launched the experiment. The experiment began with the presentation of the attention getter, a colorful star jumping against a black background on the center of the screen. The visual stimulus was paired with the repetitions of two tokens of the word "Kan (look)". One token was 441ms and the other was 412ms. The two tokens were clustered in a random order with a 1000ms pause in between. As soon as the infant looked at the center of the screen, the experimenter turned on the pretest trial, a cat zooming in and out against a grey background on the center. The speech accompanying the visual stimulus was comprised of three sentences – "Zhe Shi Shenme? (What's this?) Mao (cat), Mao, Mao; Zhe Shi Mao (This is a cat), Mao, Mao; Yi Zhi Mao (a cat), Mao, Mao". The pretest trial served to familiarize the infant with the procedure. After the pretest trial, the habituation phase began immediately. The infant was presented with trials containing one tone, either the rising tone or the dipping tone. Seven tokens of the tone were presented randomly and repeatedly until the end of the habituation phase. The inter-stimulus-interval was 1000ms. Each trial was initiated upon infant's looking, and was terminated if he or she

looked away for at least two seconds or if the maximum trial length （21s） was reached. When the looking time of three consecutive trials declined to 50% of the total of the first three habituation trials, the experiment proceeded into the test phase automatically.

During the test phase, two trial types were presented, Same type versus Different type. The Same test trial presented 6 novel tokens of the same tonal category of the habituation phase. The Different test trial presented 6 tokens of the other tone – the contrasting tone. The order of the presentation of the two trial types was counter-balanced across infants. The inter-stimulus-interval was 1000ms, and the maximum trial length was 21s. The visual stimulus used was the same as the habituation phase. Following the test phase there was one post-test trial, which presented the cat picture as in the pretest trial, accompanied by the repeated auditory stimulus of "Mao (cat) ". The post-test was distinct from the habituation and the test trials, and infants should increase looking during this trial relative to the last test trial. This allowed us to determine whether infants were still engaged in the task at the end of the test phase. For each trial, if the infant looked away from the screen for at least two seconds or if the maximum trial length elapsed, the trial would end, and the attention getter popped up automatically to attract the infant's attention back to the screen. Once the infant gazed at the screen, the experimenter pressed a computer key to initiate a trial. While infants were divided into the rising tone group and the dipping tone group for the habituation phase, all infants heard the same test trials.

## 3. Results

For each infant, the looking time of the Same test trial was compared with that of the Different test trial. Because infants of this age range typically produce large variability in looking times (e.g., some infants are overall long lookers and some are overall short lookers), the transformation of logarithm to base 10 was conducted on all the data to reduce individual variability.

Our prediction was that if infants could categorically discriminate rising and dipping tones, their looking time for the Different test trial should be significantly longer than that for the Same test trial during the test phase following habituation. If infants could not categorize rising and dipping as distinct tones, looking times to the two test trials should not differ.

A mixed $2 \times 2$ analysis of variance, with the Habituation Tone (rising vs. dipping) as the between-subject factor and Test Tone Type (Same or Different) as the within-subject factor, revealed a significant main effect of Test Tone Type, $F(1,18)=4.855$, $p=0.041$, but no main effect of Habituation Tone, $F(1,18)=0.037$, $p=0.851$ and no interaction of Habituation Tone $\times$ Test Tone Type, $F(1,18)=1.510$, $p=0.235$. These results demonstrate that infants distinguished rising and dipping tones in Mandarin Chinese.

Further analyses were conducted to compare each test trial type with the last habituation trial. Our prediction was that if infants categorically discriminated the two tones, looking time for the Different test trial should be longer than that for the last habituation trial whereas the Same test trial should not be longer in looking time relative to the last habituation trial. Paired Samples T-tests confirmed these predictions. There was no significant difference between the Same test trial and the last trial of the habituation phase, $t(19)=0.099$, $p=0.922$, 2-tailed, while the Different test trial induced significantly longer looking time than the last

habituation trial, $t(19)=2.529$, $p =0.02$, 2-tailed. See Figure 4 for the looking patterns of the last habituation and test trials.
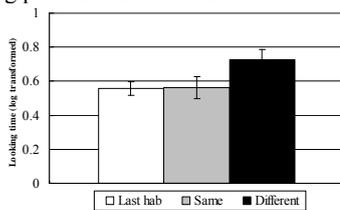


Figure 4: *Looking times (means & standard errors) of Last habituation trial, Same test trial and Different test trial.*

## 4. Discussions and conclusion

The results of the present experiment show that Mandarin-learning infants aged 4 to 13 months can discriminate rising and dipping tones in Mandarin. These results appear to contrast with Tsao's [21] finding. In Tsao's study Mandarin-learning infants around one year of age were poorer in discriminating the rising-dipping contrast than the high level versus dipping contrast, and based on the data reported in Tsao, it is unclear whether the infants discriminated the rising-dipping contrast significantly above chance. Our experiment clearly showed that infants can discriminate this tonal contrast. Note that our study and Tsao's study differ in the testing procedures, which may explain the seemingly differential results from these two studies. It may be that the habituation procedure used in our study is more sensitive in revealing infants' perceptual ability. Our results suggest that infants have no difficulty in discriminating rising and dipping tones, which are considered acoustically similar in the literature.

Furthermore, our experiment provides evidence of tonal categorization beyond simple discrimination, because the same tonal category during the test phase was a novel set of exemplars from the tonal category of the habituation tokens. Thus, the fact that infants increased looking to the Different test trial relative to the last habituation trial while yielding comparable looking to the Same test trial and the last habituation trial demonstrates that they can categorize the two similar tones. Note that categorization can also be interpreted even if infants were to increase looking for both the Same and Different test trials, as long as the increase for the Different trial is greater than for the Same trial. But the lack of increase in looking time for the Same trial that we observed in our data is stronger evidence for categorical perception of the tones.

One question remains with regards to how to define tonal similarity. What are the acoustic dimensions that determine whether certain tones are similar to one another? Pitch height, contour, onset, $\Delta F0$, average F0, duration, and pitch change inflection point are among the proposed dimensions [7, 8, 14, 17, 21, 22]. For example, the conflicting findings of the discrimination of rising and high falling tones suggest that our understanding of what constitute similar acoustic cues that are truly confusable for infants' initial perception is still limited.

In sum, our study reveals that infants from 4 to 13 months of age can categorize the two similar tones – rising and dipping, before full-blown speech production. But because the age range in our study is large, it is unknown whether infants can differentiate rising tone and dipping tone at birth or need to develop the perception of this tonal contrast through learning experience during the first year of life. We plan to test more infants at different ages to further understand the developmental time course for the acquisition of tones.

## 5. References

[1] Best, C. T., *et al.*, "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants", *J. of Experimental Psychology: Human Perception and Performance,* 14(3): 345-360, 1988.

[2] Chao, Y. R., *Mandarin Primer*. Cambridge: Harvard University Press, 1948.

[3] Gauthier, B., R. Shi, and Y. Xu, "Learning phonetic categories by tracking movements", *Cognition*, 103(1): 80-106, 2007.

[4] He, J. and Y. Wang, "Lun Duiwai Hanyu Yuyan Jiaoxuezhong Shengdiaode Jiupian", *Beijing Lianhe Daxue Xuebao,* 14(2): 2000.

[5] Karzon, R. G. and J. G. Nicholas, "Syllabic pitch perception in 2- to 3-month-old infants", *Perception & Psychophysics,* 45(1): 10-14, 1989.

[6] Kuhl, P. K., *et al.*, "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months", *Developmental Science*, 9(2): F13-F21, 2006.

[7] Lee, K. Y. S., *et al.*, "Tone perception ability of Cantonese-speaking children", *Language and Speech*, 45(4): 387-406, 2002.

[8] Lei, M. K. -Y., "Discrimination of level tones in Cantonese-learning infants", in *ICPhS*, Saarbrücken, 2007.

[9] Li, C. N. and S. A. Thompson, "The acquisition of tone in Mandarin-speaking children", *J. of Child Language,* 4(2): 185-199, 1977.

[10] Liu, C. Y., *18-23 ge yue Ertong Putonghuade Yuyin Fazhan (Shanghai Diqu)* , M.A., Shanghai Normal University, Shanghai, 2004.

[11] Liu, S. S., *Zhongji Hanyu Shuiping Hanguo Liuxuesheng Shengdiao Shengxue Shiyan Yu Fenxi*, M.A., Guangxi Normal University, Guangxi, 2005.

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

[13] Mattock, K., *et al.*, "The developmental course of lexical tone perception in the first year of life", *Cognition*, 106(3): 1367-1381, 2008.

[14] Moore, C. B. and A. Jongman, "Speaker normalization in the perception of Mandarin Chinese tones", *J. of the Acoustical Society of America*, 102(3): 1864-1877, 1997.

[15] Narayan, C., *et al.*, "The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination", *Developmental Science,* in press.

[16] Polka, L., *et al.*, "A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern", *J. of the Acoustical Society of America,* 109(5): 2190, 2001.

[17] Shen, X. S. and M. Lin, "A perceptual study of Mandarin tones 2 and 3", *Language and speech*, 34(2): 145-156, 1991.

[18] Shi, R., "Contextual variability and infants' perception of tonal categories", *Chinese Journal of Phonetics*, in press.

[19] So, C. K., "Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences", in *the 11th Australian ICSST*, University of Auckland, New Zealand, 2006.

[20] Sundara, M., *et al.*, "Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English", *Cognition,* 100(2): 369-388, 2006.

[21] Tsao, F.-M., "The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants", *Chinese Journal of Psychology*, 50(2): 111-124, 2008.

[22] Tsao, F.-M., *et al.*, "Different trends of developing lexical-tone perception in Mandarin-learning infants", presented at the SRCD 2007 Biennial Meeting, Boston, USA, 2007.

[23] Wang, A. H., "Hanyu Shengdiao Tezheng Jiaoxue Tantao", *Yuyan Jiaoxue Yu Yanjiu*, (3): 70-75, 2006.

[24] Wong, P., *et al.*, "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children", *J. of Speech, Language, and Hearing Research*, 48(5): 1065-1079, 2005.

[25] Yang, J., *Putonghua Ertong Shengdiao Huode Yanjiu*, M.A., Hunan University, Changsha, 2006.