A Reduced Sensitivity to Tones in Young Tone Learners’ Word Recognition

Weiyi Ma, Liqun Gao, and Peng Zhou

1. Introduction

Unlike most Indo-European languages, tone languages rely heavily on lexical tones (thereafter, tones) to distinguish word identity (Hu, Gao, Ma, & Yao, 2012). For example, Mandarin, the most widely spoken tone language, uses four tones to define word identity. Therefore, ma1 with a high-level tone means mother; ma2 with a rising tone means hemp; ma3 with a dipping tone means horse; and ma4 with a falling tone means to curse. Despite the fact that approximately 70% of the world’s languages are tone languages, tone acquisition still remains understudied (Yip, 2002). The present study examines 3-year-old Mandarin speakers’ ability to use tones in word learning.

Tone sensitivity emerges early in infancy (Cheng et al., 2013). A recent study revealed tone discrimination in Mandarin- and Cantonese-reared infants as young as 4 months of age (Yeung, Chen, & Werker, 2013). Research also showed that 9-month-old Mandarin-reared infants could discriminate Thai tones, whereas their English-reared counterparts failed to show such discrimination, suggesting that tone perception is shaped by language exposure (Mattock & Burnham, 2006).

How do Mandarin-learning children acquire tones? Research has shown that Mandarin-English bilingual infants use tones to distinguish Mandarin words at 11 months of age (Singh & Foong, 2012). Wong, Schwartz, and Jenkin (2005) examined tone perception in Mandarin-speaking 3-year-olds, who were presented with tone minimal pairs (e.g., tang1 [soup] vs. tang2 [candy]) on a picture-name matching task. Children’s accuracy rate was as low as 69% when T3 was tested though they reliably perceived other tones. The later age of acquisition (AoA) of T3 might be related to the fact that T3 was perceptually
less distinctive than other tones (So & Best, 2010). However, these studies used familiar words, making it almost impossible to evaluate children’s ability to use tones to distinguish word identity independent of their lexical knowledge of Mandarin words.

Using the Intermodal Preferential Looking Paradigm (IPLP – Golinkoff, Ma, Song, Hirsh-Pasek, 2013), Singh, Hui, Chan, and Golinkoff (2014) recruited 18- and 24-month-old Mandarin-English bilingual children. Children were first familiarized with two novel objects, each paired with a novel label (e.g., leng2). The two objects were then presented side-by-side accompanied by speech stimuli directing children to look at one of the objects. The target word was either correctly pronounced (CP: leng2) or mispronounced with tone change (MP-tone: leng4) or vowel change (MP-vowel: ling2). In the Mispronunciation Paradigm, sensitivity to tone and vowel change was indicated by a significant reduction in visual fixation on target images on MP trials compared to CP trials (Mani & Plunkett, 2007; Swingley & Aslin, 2000; White & Morgan, 2008). Both age groups found the target images on CP trials but not on either MP-tone or MP-vowel trials, suggesting that children were equivalently sensitive to vowels and tones in word recognition.

Using an experimental design similar to Singh et al.’s study (2014), a recent study examined the influence of tone and vowel change on monolingual Mandarin-speaking 2- and 3-year-olds’ learning of new words (Ma, Zhou, Singh, & Gao, 2017). The 3-year-old children found the target images when the target word was correctly pronounced or mispronounced with tone change. Thus, it appears that tones may not be reliable cues to word identity for Mandarin-speaking 3-year-olds in learning of novel words. However, the finding is inconsistent with the fact that Mandarin acquisition requires the use of tones as a distinctive feature.

Why did the children not use tones to learn novel words in Ma et al.’s study (2017)? There are two possible explanations. First, the finding may be related to the experimental procedure – the mispronunciation paradigm – which tested tone knowledge on the MP trial where neither of the two images children saw matched the mispronounced word precisely. Thus, this design, a two-alternative forced choice task, might have encouraged children to map the tone change to the image that was more likely to be the target. This choice, however, did not necessarily mean that children did not understand the function of tones in Mandarin. Second, presumably due to the diversity of the functions of pitch (or tone) variations (e.g., affect and emphatic stress –Birch & Clifton, 2002), cues to talker identity –Remez, Fellowes, & Rubin, 1997, question/statement distinction –van Heuver & Haan, 2002), tones have a less than transparent relation to word identity even in tone languages. Thus, learners may need additional assistance to map tones to new words, for example, an exposure to the use of tones in distinguishing minimal pairs of words.

Therefore, it is unclear whether Mandarin-speaking 3-year-olds can use tones to distinguish novel words under experimental conditions. In addition, since Ma et al. (2017) manipulated only T2 and T4, it is uncertain whether other
tone pairs differ in their likelihood to be used as cues to word identity. For example, is T3 less likely than other tones to be used as cues to word identity given its later AoA and lower level of perceptual distinctiveness relative to other tones (So & Best, 2010)?

2. Present study

This study examines whether Mandarin-speaking 3-year-olds can map tones to labels when a context indicating the function of Mandarin tones is provided, and when their tone knowledge is tested with an image precisely matching the target word. In addition, it examines whether T3 is more difficult to map to words than other tones.

2.1. Participants

Twenty-six 3-year-old ($M = 37.89$ months, range = 31.45–41.78 months; 13 boys) monolingual Mandarin-speaking children participated.

2.2. Apparatus and stimuli

Children sat on a blindfold female research assistant’s lap facing a 39 in LED TV monitor at a distance of 1 meter from the center of the screen. The visual stimuli were showed either in the center (e.g., during the training) or as left and right split-screen displays (e.g., during the test) at children’s eye level; the auditory stimuli were presented through the TV monitor. A hidden camera recorded children’s visual fixation to the display for offline coding.

Images of four novel objects were used. A female native speaker of Beijing Mandarin produced the speech stimuli. To maintain children’s interest in the task, she produced the speech in a child-directed manner (Cooper & Aslin, 1990; Fernald, 1985; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Werker, Pegg, & McLeod, 1994). Two made-up word, /ʃɔ/ and /tiu/, were used, each pronounced in three versions with T2 (rising), T3 (dipping), and T4 (falling) respectively.

2.3. Procedure

This experiment used the IPLP. An experiment consisted of a T2/4 block (a T2 word and a T4 word) and a T2/3 block (a T2 word and a T3 word). The order of presentation of the test blocks was counterbalanced. Each block consisted of four phases. During the salience phase (one trial), participants were shown two novel objects side-by-side without animation or accompanying auditory stimuli. During the training phase (four trials, 24 seconds each), participants were presented with the animated videos of the same two novel objects as in the salience phase, each paired with one of the two training sentences, twice each in alternating order (e.g., /ʃɔ2/-/ʃɔ4/ in the T2/4 block; /tiu2/-/tiu3/ in the T2/3 block). Words were randomly assigned to appear in the T2/4 block or the T2/3
block. The pairing of the image and labels was counterbalanced. During the test phase, participants were presented with the static version of the two novel objects side-by-side and were directed to look at one of them. In each test block, there were four test trials, two for each word. Thus, there were two T2 trials and two T4 trials in the T2/4 block, and two T2 trials and two T3 trials in the T2/3 block (see Ma, W., Zhou, P., Singh, L., & Gao, L. (2017) for details).

2.4. Coding and data analysis

Using SuperCoder, participants’ eye movements were coded frame-by-frame to the thirtieth of a second. Two dependent variables were used.

Latency: Duration of the time, starting from 367 msec after the onset of the target word to the time when participants launched their initial fixation on the target image.

Proportion of total look to the target (Target fixation): Target fixation was calculated by dividing a participant’s looking time to target by looking time to the distractor and target during a trial.

2.5. Results and discussion

2.5.1. Latency

We collapsed latencies across trials to generate an average latency for each word within a block. Separate paired-samples t tests demonstrated that in the T2/4 block, latencies did not differ between T2 (M = .21 sec, SD = .21) and T4 words (M = .37 sec, SD = .38; p = .12). However, in the T2/3 block, latencies were shorter for T2 words (M = .24 sec, SD = .41) than for T3 words (M = 1.18 sec, SD = .82; t(15) = 4.15, p = .001, Cohen’s d = 1.04). Thus, only T3 words were associated with a processing cost with respect to latency.

2.5.2. Target fixation

Target fixation was compared to chance (0.5) to evaluate children’s word-learning performance. Successful word learning was indicated by the target fixation, which was significantly above chance. In the T2/4 block, target fixation differed from chance on both T2 (t(25) = 5.07, p = .0001, Cohen’s d = 1.99)) and T4 trials (t(25) = 2.97, p = .006, Cohen’s d = 1.16), suggesting that when children were presented with T3 and T4 words, they successfully learned both words. In the T2/3 block, target fixation differed from chance on T2 trials (t(25) = 5.35, p = .0001, Cohen’s d = 2.01) but not on T3 trials (p = 0.36). Thus, when children were presented with T2 and T3 words, they learned the T2 word but failed to learn the T3 word. Children’s performance on the T3 tests was further examined in the following analyses.

Did children look at the target image and the distractor randomly on the T3 test trials? The 26 children were divided into two groups, high performers (n=13) and low performers, based on the median split of their target fixation on the T3
tests (0.45). Separate one sample t tests compared target fixation to chance (0.5) within each group. The high performers looked at the target image more than the distractor (t(12) = 3.08, p < 0.001, Cohen’s d = 1.71). Thus, they successfully mapped the T3 label to the target image. However, the low performers looked at the distractor more than the target image (t(12) = 10.17, p < .001, Cohen’s d = 5.64), suggesting that they consistently mismapped the T3 label to the image of the T2 word.

Was the low performers’ misuse of T3 as T2 driven by a factor unrelated to their representation of T3? The high and low performers did not differ in their actual age (p = 0.83), attention in the training (p = 0.87) or test (p = 0.81) phases, or latency on the T3 tests (p = 0.11) in the T2/3 block. In addition, in the T2/3 block, the low performers did not show preference for a particular image (p = 0.73) or for a particular label in the training as indicated by the fact that their average attention across the training trials did not differ between the T2 and T3 words (p = 0.43). Furthermore, their latency of the T3 trials (M = 1.22 sec, SD = 1.10) was longer than that of the T2 trials (M = 1.22, SD = 1.10) in the T2/3 block (t(12) = 2.57, p < 0.05, Cohen’s d = 0.71), suggesting that they distinguished between the T2 and T3 words in the test. A one sample t test examined the low performers’ word learning performance in the T2/4 block and found that their target fixation was significantly above chance (M = 0.62, SD = 0.13, t(12) = 3.32, p < 0.001, Cohen’s d = 1.84) and an independent samples t test showed that their target fixation did not differ from the high performers in the T2/4 block (M = 0.63, SD = 0.11, p = 0.77), suggesting that the low performers did not have difficulty with the experimental paradigm and they were not weak word learners overall. Similar to the analysis on the T3 test, we equally divided children into two groups on each of the other tests respectively, based on the median splits of target fixation on that particular test. We did not find a contrasting pattern of word learning performance between groups similar to the T3 test on any other tests. Thus, the low performers’ misuse of T3 as T2 on the T3 test appeared to be driven by their representation of T3.

This study showed that 3-year-old children could map the tone pair of T2 and T4 to labels; however, when presented with the tone pair of T2 and T3, children as a group could map T2 but not T3 to a label. Furthermore, the high performers learned the T3 words, but the low performers consistently misused T3 as T2. In addition, children found the target image least efficiently when learning T3 words relative to learning T2 and T4 words. Thus, Mandarin-speaking 3-year-olds can use tones to learn novel words, and that children have more difficulty using T3 than other tones as a cue to word identity in a word learning task. When children have an inaccurate representation of T3, they tend to use T3 as T2 in learning novel words.

There are two explanations to the low performers’ use T3 as T2. First, perhaps children learned the T2 word because it was highly distinctive from other tones; however, the perceptual distinction between T2 and T3 was not salient enough to enable the low performers to use T3 as a cue to word identity despite their detection of the tone difference. Second, it may be related to the
Mandarin tone sandhi effect. In Mandarin, when T3 words occur in combination, the first T3 becomes rising – T2 (Yip, 2002). Thus, Mandarin-speaking children’s language input might bias them towards the tone modification from T3 to T2 but not vice versa.

This study showed that tones might not necessarily be reliable cues to word identity even for tone speakers. This may be related to the fact that for both tone and non-tone speakers, pitch variations also convey information of affect and emphatic stress (Birch & Clifton, 2002), cues to talker identity (Remez et al., 1997, and question/statement distinction (van Heuver & Haan, 2002). Thus, tones may have an ambiguous relation to word identity even for tone speakers. A recent study showed a similar pattern of results in Mandarin-English bilingual children (Singh, Goh, & Wewalaarachchi, 2015).

The weak lexical link between tone and meaning may be also related to the origin and evolution of speech and music. Charles Darwin (1871) hypothesized that speech and music originated from a common emotional signalling system based on the imitation and modification of sounds in nature. This hypothesis is supported by recent evidence that speech and music share underlying cognitive and neural resources (Yang, Ma, Gong, Hu, & Yao, 2014) and draw on a common code of prosodic attributes when used to communicate emotional states (Thompson, Marin, & Stewart, 2012). Changes in frequency spectrum, intensity, and tempo that evoke emotional responses in music and speech also trigger emotions when perceived in nonlinguistic, non-musical environmental sounds (Ma & Thompson, 2015). Therefore, the referential function of tones is not unique to language, rendering the relation between tones and word perception less than transparent.

References


Golinkoff, Roberta, Ma, Weiyi, Song, Lulu, & Hirsh-Pasek, Kathy. 2013. Twenty-five years using the Intermodal Preferential Looking Paradigm to study language acquisition: What have we learned? Perspectives on Psychological Science, 8, 316-339.


Yang, Hua, Ma, Weiyi, Gong, Diankun, Hu, Jiehui, & Yao, Dezhong. 2014. A longitudinal study on children’s music training experience and academic development. *Scientific Reports,* 4, 5854.

