Three-year-old tone language learners are tolerant of tone mispronunciations spoken with familiar and novel tones

Weiyi Ma and Peng Zhou

Abstract: An important issue in language acquisition is understanding the function of suprasegmental information (e.g., tones) in spoken word recognition. Recent research found that three-year-old monolingual Mandarin learners recognized Mandarin words that were mispronounced using another Mandarin tone. This finding suggests that tone learners may have a tolerance of tone variation in spoken word recognition (i.e., their acceptance of tonally mispronounced words) by three years of age. However, it is also possible that the three-year-olds’ tolerance of tone variation in spoken word recognition was merely driven by the children’s familiarity with Mandarin tones, since the tone mispronunciations were produced using another Mandarin tone. This study examined whether three-year-old Mandarin learners can recognize words mispronounced using a novel tone that is unrelated to the Mandarin tone system, and whether their word recognition performance differs among words that are correctly pronounced, mispronounced with another Mandarin tone, and mispronounced with the novel tone. Two major findings emerged. First, spoken word recognition was faster on correctly pronounced words than on
mispronounced words. Second, the children recognized the tested words regardless of pronunciation type, revealing a tolerance of tone variation in spoken word recognition. Thus, this study confirmed the recent discovery that tone acquisition exhibits a tolerance of tone variation around age three. In addition, this tolerance is not specific to tone mispronunciations produced using another Mandarin tone.

**Subjects:** Psychological Science; Developmental Psychology; Educational Psychology

**Keywords:** word recognition; language development; tones; Mandarin

In speech processing, children are exposed to various types of speech information, including segmental (consonants, vowels) and suprasegmental (e.g., stress, tone) information, which often serve different functions. For example, in English, the segments /f//l//r//b//ε//r/ produced in such an order form the word “forebear”, while the suprasegmental information (e.g., stress, tone, intonation) distinguishes its form class and conveys communicative information such as affect and emphatic stress (e.g., Birch & Clifton, 2002; Cutler, 1986). Thus, the ability to discriminate lexically relevant speech cues is crucial for language learning. Most of the research on the development of this ability has focused on children learning non-tone languages, such as English and French (e.g., Mani & Plunkett, 2007; Nazzi, 2005; Quam & Swingley, 2010; Swingley & Aslin, 2000, 2002; White & Morgan, 2008). However, approximately 70% of the world’s languages are tone languages that use both segments and lexical tones (hereafter, tones) to distinguish words (Yip, 2002), and at least half of the world’s population are tone speakers (Fromkin, 1978). Thus, the large body of evidence and theories on early language development is based on only a minority of languages and language learners, leaving tone acquisition understudied. Understanding tone acquisition is therefore essential for any complete theory of child language development.

An examination of tone acquisition speaks to a major question in language acquisition: Do tones have a weak relationship to word identity? Research on English acquisition found that compared with segments, tones have a relatively weak relationship to word identity (e.g., Quam & Swingley, 2010). However, English is a non-tone language that does not use tones to distinguish word identity, thus leaving the generalizability of the finding unclear. A major theoretical gap therefore exists in our understanding of the function of tones in defining word identity.

Do tones have a weak relationship to word identity even in tone languages? The current literature on tone language acquisition has often studied Mandarin—the most widely spoken tone language—that uses four basic tones to distinguish word identity (in addition to a fifth, neutral tone that is typically placed on weak syllables—Liu & Samuel, 2004). Thus, ma can mean mother (T1: high level), hemp (T2: rising), horse (T3: dipping), and curse (T4: falling). Tone perception is therefore crucial for Mandarin acquisition. Tone sensitivity emerges early in infancy. Infants reared in Mandarin as well as Cantonese, a Chinese language that has six tones, could discriminate tones as young as four months of age (Yeung, Chen, & Werker, 2013). Furthermore, their sensitivity to lexical stress and tones emerged earlier than that to vowels and consonants (Yeung et al., 2013). Neural responses to tone variation were observed even in newborns (Cheng et al., 2013). Using Thai tones, Mattock and Burnham (2006) found tone categorization in English- and Mandarin-reared six-month-olds, but this sensitivity only remained at nine months of age in Mandarin-reared children. A similar pattern of perceptual narrowing for tones was observed in French-reared infants (Mattock, Molnar, Polka, & Burnham, 2008), suggesting that tone sensitivity is influenced by child language environment in the first year of life. However, these finings do not speak to spoken word recognition per se.

Research also examined Mandarin-learning children’s ability of mapping tones to labels in word recognition. Singh and Foong (2012) found that Mandarin-English bilinguals start to distinguish words that contrast in tones in Mandarin at 11 months of age. A recent series of research
examined tone development in spoken word processing in 2- to 4.5-year-old Mandarin learners (e.g., Ma et al., 2017a; Ma, Zhou, Singh, & Gao, 2017b; Singh, Hui, Chan, & Golinkoff, 2014), using the mispronunciation paradigm, which is an important approach to studying children’s phonological sensitivity (Ballem & Plunkett, 2005; Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002; White & Morgan, 2008). In a typical mispronunciation paradigm, children were shown a target image and a distractor image side-by-side at the same time while the accompanying speech stimuli prompted children to look at one of them (i.e., the target image). Children’s word comprehension is measured by their differential visual fixation to the two images in this intermodal preferential looking paradigm (IPLP—Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). On some of the trials, the target word is correctly pronounced while on other trials the target word is mispronounced. This paradigm is based on the assumption that if children are sensitive to certain phonological information, mispronunciations of the information should compromise the accuracy and efficiency of word recognition (e.g., Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Havy & Nazzi, 2009; Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002; White & Morgan, 2008).

Using the mispronunciation paradigm within the IPLP, Singh et al. (2014, Experiment 2a) tested the learning of new words in 18- and 24-month-old Mandarin-English bilingual children. The children were first shown two novel objects accompanied by two novel labels (e.g., It is a leng2! [The speech stimuli were presented in Mandarin]). The two objects were then presented side-by-side in split screen displays with the accompanying speech stimuli prompting the children to look at one of them. The target word was either correctly pronounced (CP: leng2) or mispronounced with another Mandarin tone (MP: leng4). Both age groups looked at the target images more than the non-target image on CP trials but not on MP trials, suggesting that the 18- and 24-month-old Mandarin learners were highly sensitive to tones in spoken word recognition. However, novel words do not have established phonological representations, potentially rendering tone interpretation ambiguous given that tone changes may serve non-lexical functions as well. Furthermore, the developmental trajectory of tone sensitivity beyond 24 months of age still remains unclear. Research (Singh, Goh, & Wewalaarachchi, 2015) then examined three- and four-to-five-year-old Mandarin-English bilinguals’ recognition of familiar words. Using the mispronunciation paradigm within the IPLP, they found that older children had a reduced sensitivity to tones in word recognition compared to younger children.

Was the 4.5-year-old Mandarin-English bilinguals’ reduced tone sensitivity reported in Singh et al. (2015) due to their accumulating exposure to English—a none-tone language? Arguing against this possibility are the findings of a recent series of studies on tone acquisition in monolingual Mandarin-learning two and three-year-olds (Ma et al., 2017a, 2017b). Using an experimental paradigm similar to Singh et al.’s (2014, 2015), Ma and colleagues tested monolingual Mandarin-learning children’s tone sensitivity in learning new words (Exp. 1a in Ma et al., 2017b) and found that tone variation hindered word recognition accuracy in the two-year-olds who rejected the tonally mispronounced words; however, tone variation did not hinder word recognition accuracy in the three-year-olds who readily accepted the tonally mispronounced words. These findings suggested that the three-year-olds were tolerant of tone variation in spoken word recognition. This tolerance of tone variation was also observed in familiar word recognition in the three-year-olds (Exp. 2 in Ma et al., 2017b).

Thus, recent research progress revealed a tone tolerance in spoken word recognition in monolingual Mandarin acquisition at three years of age (Ma et al., 2017a, 2017b) and in Mandarin-English bilingual acquisition at four to five years of age (Singh et al., 2015), suggesting that tones may have a weak relationship to word identity even in a tone language. One possible explanation to this finding is that pitch variation serves other communicative functions beyond defining word identity in both non-tone (e.g., Banse & Scherer, 1996; Fernald & Kuhl, 1987; van Heuven & Haan, 2002) and tone languages (e.g., Liu & Pell, 2012; Yuan, 2011). Thus, this tolerance of tone variation may indicate children’s growing knowledge of the functional diversity of tones. However, it should be noted that these studies examined children’s tone sensitivity using words mispronounced with
a wrong but real Mandarin tone. This experimental design leaves one possibility open. Perhaps, the older age group (e.g., three-year-olds in Ma et al., 2017b) accepted tone mispronunciations merely because of their increasing familiarity with the Mandarin tone system, since the tone mispronunciations were produced using another Mandarin tone. Thus, it is still unclear whether the Mandarin-learning three-year-olds’ tone variation tolerance is specific to tone mispronunciations produced with another Mandarin tone (i.e., a wrong but real Mandarin tone) or is even observable when tone mispronunciations produced with a tone unrelated to the Mandarin tone system are used. An examination of this issue can verify the recent finding that tone language learners had a tone tolerance in spoken word recognition at three years of age (Ma et al., 2017a, 2017b). Furthermore, the results allow us to determine the generalizability of the tone variation variability that was originally observed using tone mispronunciations produced using another Mandarin tone.

This study examined monolingual Mandarin-learning 3-year-olds' recognition of familiar words when words were correctly pronounced, mispronounced with another Mandarin tone, and mispronounced with a novel tone unrelated to the Mandarin lexical tone system. If the 3-year-old Mandarin learners indeed have a tone variation tolerance as observed by Ma et al. (2017a, 2017b) and this tone variation tolerance is not specific to the mispronunciations produced using another Mandarin tones, the children should recognize tone mispronunciations regardless of the relevance of the tones to the mandarin tone system.

Three-year-old monolingual Mandarin learners were tested at Beijing Language and Culture University (BLCU) Kindergarten and Nursery. All children were healthy, had no history of auditory or visual impairments, and did not regularly hear any language other than Mandarin. This study used the IPLP, where children’s visual fixation time served as the dependent variable. A mispronunciation paradigm was used, a standardly used experimental paradigm in the investigation of children’s phonological sensitivity.

1. Method

1.1. Participants

The participants were 28 three-year-old children (M = 37 months, range = 31–40 months; 14 boys). Two additional children were excluded from the final sample because of inattentiveness (n = 1) and overall side bias (n = 1, looking time on one particular side in the test was greater than 80% of the entire looking time in the test) respectively. Written informed consent was obtained from the parents of all the participants in this study.

1.2. Apparatus and stimuli

Participants were tested in a quiet testing booth in the BLCU Kindergarten and Nursery. Participants sat on a blindfolded female research assistant’s lap facing a 39 in LED TV monitor at a distance of 1 meter from the center of the screen. Visual stimuli were displayed to the left and right of the screen at eye level. Auditory stimuli were presented through internal speakers within the TV monitor. A hidden camera recorded the children’s visual fixation to the display. Video recordings were then coded offline.

Prior to the child study, a female native speaker of Beijing Mandarin produced auditory stimuli in a sound-attenuated recording chamber. Speech stimuli were produced in a child-directed manner (Cooper & Aslin, 1990; Fernald, 1985; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Werker, Pegg, & McLeod, 1994). This study used 12 words (ji1 [chicken]-che1 [car]; cao3 [grass]-shou3 [hand]; xie2 [shoe]-qiu2 [ball], bi3 [pen]-gou3 [dog], bao1 [bag]-zhu1 [pig], and cao4 [vegetable]-shu4 [tree]). These words were selected based on five criteria: 1) they are familiar nouns for Mandarin-speaking 3-year-olds based on parental report on the MacArthur Communicative Developmental Inventories (Tardif, Fletcher, Zhang, Liang, & Zuo, 2008); 2) they do not begin with certain nasal, lateral, or approximant consonants (i.e.,/j/,/l/,/m/,/n/), which might modulate the pitch contours (Clark, Yallop, & Fletcher, 2007); 3) they are open syllable words, since the coda might adapt the pitch contour...
(Clark et al., 2007); 4) they are related to images that are readily presentable through visual displays; and 5) words paired together had the same tone, which ensured that any tone variation used on MP trials had the same tonal distance from the two words. As a result of these constraints, only twelve words were used in this study. The visual stimuli were 12 images for each of the 12 words. To maintain the children’s attention, seven slightly different carrier phrases were used, in which the target words occurred in utterance final position (Table 1). Some of the sentence frames used the generic classifier (i.e., ge4), which is not indicative of the target. Target words were randomly paired with carrier phrases and then synthesized into a unit using Audacity 2.0.3.

To test children’s recognition of words mispronounced using a novel tone, we created a rising-falling tone by inverting the tone contour of T3 (Figure 1). This novel tone was used because it does not distinguish word identity in Mandarin and it sounds clearly distinct from the four Mandarin tones based on the native speakers of Mandarin who designed this study. Each word was either correctly pronounced, or mispronounced with another Mandarin tone (which is also a real Mandarin word but does not refer to either object on the screen), or mispronounced with the novel tone. Separate one-way ANOVAs showed that the duration, intensity, and average pitch height of the words did not differ among correct pronunciations (CP), mispronunciations with another Mandarin tone (MP-Mandarin), and mispronunciations with the novel tone (MP-novel) (Table 2).

To validate our stimuli, 15 adult native speakers of Mandarin were administered an auditory perception task, in which they were asked to identify the tone of the words by marking T1, T2, T3, T4, or unknown. All participants correctly identified the tones of the all the words used in this study, verifying the tone assignment of the speech stimuli used in this study.

Another group of native speakers of Mandarin (n = 25) rated the familiarity of the words used in this study. On each trial, the participant heard one word and was asked to rate its familiarity (1 = not familiar, 7 = highly familiar) as quickly and accurately as possible based on their knowledge of Mandarin words. Average familiarity ratings were calculated for each type of pronunciations. Paired-sample t tests then compared the average ratings across pronunciation types. Results showed that the CP (M = 5.18, SD = 1.19) and MP-Mandarin (M = 5.25, SD = 1.05) words did not differ in familiarity ratings (p = .46). However, the MP-novel words were rated as less familiar.

### Table 1. Carrier phrases used in this study

| Carrier phrases | Character by character English translation | Duration (seconds) |
|-----------------|--------------------------------------------|--------------------|
| 1 Kan4! Na3 yi 2 ge4 shi 4 | Look! Which one classifier* is | 2.31 |
| 2 Ni2 qiao2! Na3 er shi4 | You look! Which particle is | 2.37 |
| 3 Kuai4 qiao2! Na3 er shi4 | Quickly look! Which particle is | 2.35 |
| 4 Kan4 kan4! Na3 ge4 shi4 | Look! Which classifier is | 2.17 |
| 5 Kuai4 kan4! Na3 er shi4 | Quickly look! Which particle is | 2.25 |
| 6 Qiao2! Na3 yi 2 ge4 shi 4 | Look! Which one classifier is | 2.34 |
| 7 Ni2 kan4! Na3 ge4 shi4 | You look! Which classifier is | 2.14 |

*Functional words are italicized.
than both the CP (t(24) = 10.08, p < .001, Cohen’s d = 2.02) and MP-Mandarin words (t(24) = 10.77, p < .001, Cohen’s d = 2.15).

Finally, a third group of native speakers of Mandarin (n = 25) rated the perceptual naturalness of the synthesized sentences used in this study. On each trial (28 trials in total), the participant heard one sentence and was asked to rate the perceptual naturalness (1 = not natural, 7 = highly natural) of the transition between the sentence frame and the last word. Again, the average familiarity ratings were calculated for each type of pronunciations. Paired-sample t tests then compared the average ratings across pronunciation types. Results showed that perceptual
naturalness ratings did not differ between the CP (M = 4.75, SD = 1.09), MP-Mandarin (M = 4.70, SD = 1.05), and MP-novel trials (M = 4.57, SD = 1.00, p’s > .39).

1.3. Procedure
The experimental procedure was almost identical to that of Singh et al. (2015) and Ma et al. (2017b – Exp. 2)—a typical procedure used in investigating children’s phonological sensitivities in familiar word recognition (e.g., Mani & Plunkett, 2007). In this mispronunciation paradigm, children saw two images side-by-side for six seconds on each trial, while the onset of the target word began 2633 ms into a trial (e.g., Look! Where is the [target word]?). Since children’s responses to the target word were calculated starting from 367 ms after the onset of the target word, this onset divided a trial into two 3-second phases: a pre-target phase and a post-target phase. A significant increase in looking time at the target image across phases indicated understanding of the target word. Removal of responses faster than 367 ms factors in the time taken to launch an eye movement in response to auditory input, which is a typical experimental procedure and data analysis protocols used in the investigation of children’s phonological knowledge (e.g., Ma et al., 2017a, 2017b; Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002).

An experiment consisted of 28 trials. Words with the same tone were paired together (ji1 [chicken]–che1 [car]; cao3 [grass]–shou3 [hand]; xie2 [shoe]–qiu2 [ball]; bi3 [pen]–gou3 [dog], bao1 [bag]–zhu1 [pig], and cai4 [vegetable]–shu4 [tree]), which ensured that any tone variation had the same tonal distance from the two words. Pictures were yoked in pairs throughout an experimental session for one individual participant. None of the pictures were repeated on consecutive trials. An experiment contained 12 correct pronunciation (CP) trials, 10 MP-Mandarin trials, and six MP-novel trials. The number of trials was determined based on the following design considerations. First, we did not use an experimental design where the number of trials was matched across the three pronunciation types because in such an experiment, children would hear twice as many mispronunciation trials (with the MP-Mandarin and MP-novel trials combined) as correct pronunciation trials. Second, we did not match the number of trials across the two types of mispronunciation because the MP-Mandarin trials were related to more possible pairs of tone changes than the MP-novel trials, since each word may be tonally mispronounced in three different ways (using any of the three wrong tones) on an MP-Mandarin trial but in just one way (using the novel tone) on an MP-novel trial. Thus, the current experimental design was used because 1) it was not predominantly biased towards mispronunciation trials and 2) it allows us to test multiple possible pairs of tone change on the MP-Mandarin trials.

Thus, each experiment contained 12 CP trials, each testing one of the 12 words. In addition, there were 10 MP-Mandarin trials, where no target words occurred twice. The 10 MP-Mandarin trials contained all six possible pairs of tone variation for four tones regardless of the order: T1 was mispronounced with T2, T3, and T4; T2 was mispronounced with T3 and T4; T3 was mispronounced with T2 and T4; T4 was mispronounced with T1. Given that this study only had 10 MP-Mandarin trials, we did not test all possible pairs of tone variation (if the order of the two tones in a pair was taken into consideration). Finally, there were six MP-novel trials, where six of the target words were randomly selected and pronounced with the novel tone. These selected words were used for all participants. According to the procedure used in the previous research (Ma et al., 2017b; Swingley & Aslin, 2000), four stimulus orders were created. The second order mirrored the left/right assignment of the target images of the first; the third and fourth orders reversed the trial order of the first and second orders. Within each stimulus order, no more than two trials of the same pronunciation type (CP, MP-tone, MP-novel) were presented consecutively.

1.4. Coding and data analysis
Using SuperCoder (Hollich, 2005), participants’ eye movements were coded frame-by-frame to 1/50 of a second with the audio turned off so that the coder was blind to condition. Recording of 20% of the subjects by another coder yielded an inter-coder agreement of 99%.
Based on the established procedure (Mani & Plunkett, 2007), data analysis only included the trials in which children had an attention span of more than 20% in both the pre-target and post-target phases (attention during a phase was less than 20% of the length of the phase) (Quam & Swingley, 2010; Singh et al., 2014) and during which the children fixated on both the target and the distractor in the pre-target phase (Mani & Plunkett, 2007). Based on these criteria, we excluded 25 trials across all participants. As before, we examined latency and target fixation.

Two dependent variables were used. First, latency (response latency to initiate a shift from the distractor to the target—word recognition efficiency) was calculated starting from 367 ms after the onset of the target word to the time when participants launched their initial target fixation. Based on previously established procedures (e.g., Mani & Plunkett, 2007; Swingley & Aslin, 2000), only the cases where participants’ initial visual fixations were directed to the distractor were analyzed. Second, proportion of time spent fixating the target (target fixation time) was calculated by dividing the length of looking time to the target image by the total length of looking time to the target and distractor images during a test trial. Target fixations of the pre-target and post-target phases were calculated and compared. Response latency and target fixation are measures of efficiency and accuracy of word identification.

2. Results

2.1. Latency

Did latency differ among pronunciation types? Only the trials where the participants’ initial visual fixations were directed to the distractor were analyzed. There were three children who did not have latency data for the MP-novel trials because they directed their initial visual fixations to the target on all MP-novel trials. In order to compare latency across the three pronunciation types, the three children were excluded from the relevant data analyses. The final dataset contained 169 CP trials (50.3% of the 336 CP trials), 137 MP-Mandarin trials (48.9% of the 280 MP-Mandarin trials), and 80 MP-Novel trials (53.3% of the 150 MP-novel trials [only 25 participants were analyzed]). For each child, we calculated average latencies for the CP, MP-Mandarin, and MP-novel trials respectively.

A one-sample repeated measures ANOVA on latency revealed a main effect of pronunciation type ($F(2,48) = 5.76, p = .006, \eta^2_p = .19$), showing that latency differed across pronunciation types. Separate paired-sample $t$ tests then compared latency across pronunciation types. Results showed that latency on the CP trials ($M = .49$ sec, $SD = .23$) was shorter than that on the MP-Mandarin trials ($M = .84$ sec, $SD = .44$; $t(27) = 3.68, p = .001$, Cohen’s $d = .70$) and than that on the MP-novel trials ($M = .88$ sec, $SD = .51$; $t(24) = 2.88, p = .008$, Cohen’s $d = .58$) (Figure 2). However, latency did not differ between the MP-Mandarin and MP-novel trials ($t(24) = .68, p = .50$). These findings suggested that word recognition was faster with CP words than with MP words. In addition, the speed of word recognition did not differ between the two types of tone mispronunciations.

To determine if the latency data reported above were merely driven by a small subset of participants who might have directed their initial visual fixations predominately to the distractor, we calculated the number of latency trials for each participant. For the CP trials, on average, the

Figure 2. Latency. Children oriented faster to target images on CP than on MP-Mandarin and MP-novel trials. Latency on MP-Mandarin and MP-novel trials did not differ from each other. Error bars reflect SEM.
participants directed their initial visual fixations to the distractor on 6.04 trials (SD = .88, range = 5–8). There was only one participant who was two standard deviations (2SD) away from the average by having eight latency trials. For the MP-Mandarin trials, on average, the participants directed their initial visual fixations to the distractor on 4.89 trials (SD = 1.52; range = 2–8, those who had two latency trials (n = 2), those who had eight latency trials (n = 1)). There was only one participant who was 2SD away from the average by having eight latency trials. For the MP-novel trials, among the 25 participants whose latency data were analyzed and reported above, the participants directed their initial visual fixations to the distractor on 3.04 trials (SD = 1.16, range = 1–4, those who had one latency trial (n = 2), those who had four latency trials (n = 9)). Furthermore, there were no participants who were 2SD away from the average. Thus, it is highly unlikely that the latency data reported above were be driven by a small subset of the participants.

Did latency change as the experiment progressed? We divided the trials of each pronunciation type into two equally sized subsets—the earlier trials and the later trials—based on the order of presentation for each participant. Thus, each participant had six earlier CP trials, five earlier MP-Mandarin trials, three earlier MP-novel trials, and the same numbers of later trials for each pronunciation type. For the MP-Mandarin trials, there were two participants (among the 28 participants) who did not have latency data for the earlier trials. For the MP-novel trials, among the 25 participants whose latency data were analyzed and reported above, there were two participants who did not have latency data for the earlier trials and one participant who did not have latency data for the later trials. These children were excluded from the relevant data analyses. Then, for each participant, we calculated the average latency for the earlier and later trials respectively within each pronunciation type. A 2 (Time: earlier, later) x 3 (Pronunciation Type: CP, MP-Mandarin, MP-novel) repeated measures ANOVA on latency revealed only a significant main effect of Pronunciation Type (F(2,38) = 4.06, p = .025, η² = .18), suggesting that latency differed across pronunciation types. However, neither the main effect of Time (F(1,19) = 2.91, p = .11) nor the Time x Pronunciation Type interaction (F(2,38) = 2.52, p = .09) approached significance, suggesting that latency did not change as the experiment progressed (Table 3).

### 2.2. Target fixation (proportion of time spent fixating the target)

Did children recognize the words under the three pronunciation conditions? For each participant, we calculated average target fixation for the pre-target and post-target phases respectively within each pronunciation type. A 2 (Phase: pre-target, post-target) x 3 (Pronunciation Type: CP, MP-Mandarin, MP-novel) repeated measures ANOVA on target fixation revealed significant main effects of Phase (F(1,27) = 111.08, p < .001, η² = .80) and Pronunciation Type (F(2,54) = 12.76, p < .001, η² = .32) and a marginally significant Phase x Pronunciation Type interaction (F(2,46) = 2.77, p = .07). Then, separate paired-sample t tests compared the pre- and post-target target fixations within each pronunciation type. A significant increase in target fixation across phases indicates that the participant has mapped the verbal label onto the visual target. Results showed a significant increase in target fixation across phases on the CP (t(27) = 11.17, p < .001, Cohen’s d = 2.11), MP-Mandarin (t(27) = 7.02, p < .001, Cohen’s d = 1.33), and MP-novel trials (t(27) = 4.74, p < .001, Cohen’s d = .90) respectively, suggesting that children found the target images regardless of pronunciation types (Figure 3, Table 4).

|                  | Earlier trials (Mean and SD) | Later trials (Mean and SD) |
|------------------|-----------------------------|---------------------------|
| CP               | .55 sec (± .31)             | .42 sec (± .33)           |
| MP-Mandarin      | .75 sec (± .52)             | .86 sec (± .49)           |
| MP-novel         | 1.00 sec (± .65)            | .72 sec (± .58)           |

Note: Each participant had six earlier CP trials, five earlier MP-Mandarin trials, three earlier MP-novel trials, and the same numbers of later trials for each pronunciation type.
Figure 3. Target fixation in the pre- and post-target phases in Experiment 2. Target fixation significantly increased across phases on CP, MP-Mandarin, and MP-novel trials. Furthermore, the increase in target fixation on CP trials was greater than that on MP trials and MP-novel trials, but it did not differ between MP-Mandarin and MP-novel trials. Error bars reflect SEM.

Table 4. The proportion of time spent fixating the target

|          | Pre-target (Mean and SD) | Post-target (Mean and SD) |
|----------|--------------------------|---------------------------|
| CP       | .48 (± .06)              | .73 (± .10)               |
| MP-Mandarin | .49 (± .07)              | .68 (± .11)               |
| MP-novel | .42 (± .12)              | .59 (± .17)               |

Note: Paired-sample t tests compared the pre- and post-target target fixations within each pronunciation type. Results showed an increase in target fixation across phases on the CP, MP-Mandarin, and MP-novel trials. In addition, we also compared the target fixation against chance (.50) in the pre-target and post-target phases respectively. Results showed that in the pre-target phase, target fixation did not differ from chance on the CP trials (t(27) = 1.43, p = .16) or the MP-Mandarin trials (t(27) = .43, p = .67), but it differed from chance on the MP-novel trials (t(27) = 3.39, p = .002, Cohen’s d = 1.28). In the post-target phase, target fixation differed from chance on the CP (t(27) = 12.36, p < .001, Cohen’s d = 4.67), MP-Mandarin (t(27) = 8.30, p < .001, Cohen’s d = 3.14), and MP-novel (t(27) = 2.73, p = .01, Cohen’s d = 1.03) trials.

Did target fixation change as the experiment progressed? Again, we divided the trials of each pronunciation type into two equally sized subsets—the earlier trials and the later trials—based on the order of presentation for each participant. We calculated the average target fixation for the earlier and later trials respectively within each pronunciation type for each participant. A 2 (Time: earlier, later) x 2 (Phase: pre-target, post-target) x 3 (Pronunciation Type: CP, MP-Mandarin, MP-novel) repeated measures ANOVA on target fixation showed significant main effects of Phase (F(1,27) = 111.74, p < .001, η² = .81) and Pronunciation Type (F(2,54) = 12.71, p < .001, η² = .32). However, neither the main effect of Time (F(1,27) = .72, p = .40) nor the interactions depending on Time approached significance (the Time x Phase interaction: F(1,27) = .40, p = .84; the Time x Pronunciation Type interaction: F(2,54) = .64, p = .53), suggesting that target fixation did not change as the experiment progressed (Table 5). Thus, the target fixation of the earlier and later trials was analyzed as combined in the following data analyses.

Table 5. The proportion of time spent fixating the target in the earlier and later trials

|          | Earlier trials | Later trials |
|----------|----------------|--------------|
|          | Pre-target (Mean and SD) | Post-target (Mean and SD) | Pre-target (Mean and SD) | Post-target (Mean and SD) |
| CP       | .48 (± .06) | .72 (± .10) | .49 (± .07) | .73 (± .10) |
| MP-Mandarin | .48 (± .08) | .68 (± .13) | .50 (± .10) | .68 (± .13) |
| MP-novel | .43 (± .13) | .58 (± .17) | .41 (± .12) | .59 (± .16) |

Note: Each participant had six earlier CP trials, five earlier MP-Mandarin trials, three earlier MP-novel trials, and the same numbers of later trials for each pronunciation type.
Was word recognition more accurate with CP words than with MP words? We examined whether target fixation increased across phases to a greater extent on the CP trials than on the MP trials. For each participant, we calculated the difference in the average target fixation between the pre- and post-target stages within each pronunciation type (post-target phase target fixation minus pre-target phase target fixation). Separate paired-sample t tests found that target fixation increased across phases to a significantly greater extent on the CP trials \((M = .25, SD = .12)\) than on the MP-Mandarin trials \((M = .18, SD = .14; t(27) = 2.72, p = .011, \text{Cohen's } d = .51)\), and to a marginally significantly greater extent than on the MP-novel trials \((M = .16, SD = .18; t(27) = 2.02, p = .054, \text{Cohen's } d = .38; \text{An alpha of } .017 \text{ was used as the cutoff for significance for three comparisons})\). However, target fixation increased across phases to a similar extent on the MP-novel and MP-Mandarin trials \((t(27) = .48, p = .64)\). These findings suggest that word recognition was more accurate on the CP trials than on the MP-Mandarin trials, and that word recognition accuracy did not differ between the two types of mispronunciation.

3. Discussion

This study examined monolingual Mandarin-learning 3-year-olds’ spoken word recognition using words correctly pronounced, mispronounced with another Mandarin tone, and mispronounced with a novel tone that is unrelated to the Mandarin word identity. The latency data showed that children recognized CP words faster than MP words, but word recognition speed did not differ between the two types of mispronunciations. The target fixation data revealed that the children successfully recognized the tested words in all the three pronunciation types, supporting the recent finding that 3-year-old monolingual Mandarin learners could be tolerant of tone variation in spoken word recognition (Ma et al., 2017a, 2017b). Furthermore, target fixation increased across phases to a greater extent on the CP trials than on the MP trials, suggesting that word recognition accuracy was better with CP words than with MP words. However, target fixation increased across phases to a similar extent on the MP-Mandarin and MP-novel trials, suggesting that word recognition accuracy did not differ between the two types of tone mispronunciations.

The current findings demonstrated that the 3-year-old Mandarin learners’ tolerance of tone variation in spoken word recognition observed by Ma et al. (2017a, 2017b) was not specific to tone mispronunciations produced with another Mandarin tone. Thus, this study verified the recent discovery that monolingual tone language learners can be tolerant of tone variation in spoken word recognition at three years of age (Ma et al., 2017a, 2017b). However, it should be noted that this tolerance of tone variation does not mean that the three-year-old mandarin learners are insensitive to tone variation, since the latency data differed between the CP and MP trials. In addition, tone mispronunciations were associated with a reduced (though still significant) increase in target fixation in spoken word recognition.

Combined with previous findings (Burnham et al., 2011; Quam & Swingley, 2010; Singh et al., 2015), this tolerance of tone variation in spoken word recognition has been observed across multiple age ranges from toddlerhood to early adolescence, across different tone inventories, in both monolingual and bilingual tone language learners, and across different tone contrasts within a tone language. There are two explanations to this tone tolerance. First, the three-year-old children already have a robust phonological representation of familiar words used in this study, which allowed them to recognize the words even when they were tonally mispronounced. In a mispronunciation paradigm—which tests tone knowledge on the MP trial where neither of the two images provide an exact match for the label—children may look at the image that is more likely to be the target.

Second, this tolerance of tone variation may be related to children’s growing knowledge of the functional diversity of pitch variation. Pitch variation also serves communicative functions beyond distinguishing word identity. For non-tone speakers, pitch variation in intonation draws attention (e.g., Fernald & Kuhl, 1987), conveys emotional information (Banse & Scherer, 1996) and emphatic stress (e.g., Birch & Clifton, 2002), and distinguishes questions from statements (e.g., van Heuven & Haan, 2002). This is also the case in tone languages (e.g., Liu & Pell, 2012; Yuan, 2011). Although intonation differs from tones, presumably due to the diversity of the functions of pitch variation (a fundamental
property of tone), tones may have a weaker relationship to word identity than segments (Quam & Swingley, 2010; Singh et al., 2015, 2014). This is supported by recent findings that pitch variation also communicates emotional meaning in music and environmental sounds (Ma & Thompson, 2015), and that infants’ early sensitivity to pitch variation in language and music was found related to each other (Chen, Stevens, & Kager, 2017; Liu & Kager, 2017). Thus, the three-year-old tone language learners’ tolerance of tone variation in spoken word recognition may indicate their acquisition of the knowledge that tones variation may not necessarily indicate a change of word identity in a tone language.

It should be noted that the current study focuses on tone perception, speaking little to the relative function between segments and tones. Research with adult tone language speakers are mixed in their conclusions on the relative role of segments and tones in constraining tone word recognition. While some studies have revealed comparable sensitivity to segments and tones (Liu & Samuel, 2007; Malins & Joanisse, 2010; Schirmer, Tang, Penney, Gunter, & Chen, 2005), others have suggested greater sensitivity to segments than to tones (Cutler & Chen, 1997; Davis, Schoknecht, Kim, & Burnham, 2016; Hu, Gao, Ma, & Yao, 2012; Repp & Lin, 1990; Taft & Chen, 1992; Tong, Francis, & Gandour, 2008; Wiener & Turnbull, 2016; Ye & Connine, 1999). Thus, the comparative weight of tones and segments is still an on-going debate, which should be further explored by future research.

Mandarin relies heavily on tones in distinguishing word identity. From this, an important question arises: Are 3-year-olds’ tone tolerances an impediment to Mandarin word learning and recognition? This is highly unlikely. Mandarin-learning 3-year-olds are aware of the function of tones in Mandarin (Ma et al., 2017b). They could use tones to learn novel words in the training when they were offered a minimal pair, differing only in tone and accompanied by two different objects (Ma et al., 2017b). Thus, their successful use of tones in word learning seems to require additional cues indicating the relevance of the function of tones in the specific task. In addition, a recent study showed that their successful use of tones in learning new words is also related to the properties of specific tones used (Burnham, Singh, Mattock, Woo, & Kalashnikova, 2017).

### 4. Conclusion

This study examined monolingual Mandarin-learning three-year-olds’ tone sensitivity in spoken word recognition. We found that the three-year-olds were tolerant of tone change in familiar word recognition when words when pronounced with another Mandarin tone or a novel tone that is irrelevant to Mandarin word identity. The finding verified that Mandarin-learning three-year-olds have a tone tolerance in spoken word recognition.

### Funding
PZ is supported by a National Social Science Foundation of China (16BYY076).

### Competing interests
The authors declare no competing interests.

### Author details
Weiyi Ma\(^1\)
E-mail: weiyima@uark.edu
ORCID ID: [http://orcid.org/0000-0002-2299-9834](http://orcid.org/0000-0002-2299-9834)

Peng Zhou\(^2\)
E-mail: zhoupeng1892@mail.tsinghua.edu.cn
ORCID ID: [http://orcid.org/0000-0002-0818-2545](http://orcid.org/0000-0002-0818-2545)

\(^1\) School of Human Environmental Sciences, University of Arkansas, Fayetteville, AR, USA.
\(^2\) Department of Foreign Languages and Literatures, Tsinghua University, Beijing, China.

### Citation information
Cite this article as: Three-year-old tone language learners are tolerant of tone mispronunciations spoken with familiar and novel tones, Weiyi Ma & Peng Zhou, Cogent Psychology (2019), 6: 1690816.

### Note
1. We also analyzed the data using only the first six CP trials, the first six MP-Mandarin trials, and all the six MP-novel trials. The data analyses, which included 168 CP trials, 168 MP-Mandarin trials, and 150 MP-novel trials (only 25 participants were included for MP-novel trials), showed a pattern of results similar to what is reported in the Results section. For the latency data, the final data dataset contained 86 CP trials (51.2% of the 168 CP trials), 77 MP-Mandarin trials (45.8% of the 168 MP-Mandarin trials), and 80 MP-Novel trials (53.3% of the 150 MP-novel trials). We calculated average latencies for CP, MP-Mandarin, and MP-novel trials respectively for each participant. Separate paired-sample t tests compared latency across pronunciation types and found that latency on CP trials (M = .55 sec, SD = .31) was marginally significantly shorter than that on MP-Mandarin trials (M = .85 sec, SD = .55; t(27) = 2.25, p = .03, Cohen’s d = .43) and than that on MP-novel trials (M = .88 sec, SD = .51; t(24) = 2.63, p = .02, Cohen’s d = .49), but latency did not differ between MP-Mandarin and MP-novel trials (t(24) = .60, p = .55; An alpha of .017 was used as the cutoff for significance for three comparisons.). For the target fixation data, we calculated average target fixation for the
pre-target and post-target phases respectively within each pronunciation type for each participant. Separate paired-sample t-tests compared the pre- and post-target target fixations within each pronunciation type. Results showed a significant increase in target fixation across phases on the CP (Pre-target: M = .48, SD = .06; Post-target: M = .72, SD = .10; t(27) = 12.22, p < .001, Cohen’s d = 2.31), MP-Mandarin (Pre-target: M = .48, SD = .08; Post-target: M = .68, SD = .13; t(27) = 6.36, p < .001, Cohen’s d = 1.20), and MP-novel trials (Pre-target: M = .42, SD = .12; Post-target: M = .59, SD = .17; t(27) = 4.74, p < .001, Cohen’s d = .90) respectively.

References

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. Cognitive Development, 17, 1125–1282. doi:10.1016/S0163-6383(02)00116-8

Bollen, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1; J. Journal of Child Language, 32, 159–173. doi:10.1017/S0142715605000465

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, 70, 614–636. doi:10.1037/0022-3514.70.3.614

Birch, S., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. Journal of Memory and Language, 47, 571–588. doi:10.1006/jml.2002.3514

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., & Kasisopha, B. (2013). Are tones phones? Journal of Experimental Child Psychology, 108, 693–712. doi:10.1016/j.jecp.2010.07.008

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2017). Constraints on tone sensitivity in novel word learning by monolingual and bilingual infants: Tone properties are more influential than tone familiarity. Frontiers in Psychology, 8, 2190. doi:10.3389/fpsyg.2017.02190

Chen, A., Stevens, C. J., & Kager, R. (2017). Pitch perception in the first year of life, a comparison of lexical tones and musical pitch. Frontiers in Psychology, 8, 297. doi:10.3389/fpsyg.2017.00297

Cheng, Y., Wu, H., Tseng, Y., Yang, M., Zhao, L., & Lee, C. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. Developmental Neuropsychology, 38, 281–300. doi:10.1080/8756564.2013.799672

Clark, J. E., Yallop, C., & Fletcher, J. (2007). An introduction to phonetics and phonology (Vol. 9). Oxford, UK: Wiley-Blackwell.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. Child Development, 61, 1584–1595. doi:10.2307/1130766

Cutler, A. (1986). Forbear is a homophone: Lexical prosoy does not constrain lexical access. Language and Speech, 29, 201–220. doi:10.1177/002383098602900302

Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. Perception & Psychophysics, 59, 165–179. doi:10.3758/BF03211886

Davis, C., Schoknecht, C., Kim, J., & Burnham, D. (2016). The time course for processing vowels and lexical tones: Reading aloud Thai words. Language and Speech, 59, 196–218. doi:10.1177/0023830915586033

Fernald, A. (1985). Four-month-olds prefer to listen to motherese. Infant Behavior and Development, 8, 181–195. doi:10.1016/S0163-6383(85)80005-9

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development, 10, 279–293. doi:10.1016/0163-6383(87)90017-8

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

Gelinkoff, R. M., Mc, S., Song, L., & Hirsh-Pasek, K. (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. doi:10.1177/1745691613484936

Hovy, M., & Nazzi, T. (2009). Better processing of consonantal over vocable information in word learning at 16 months of age. Infancy, 14, 439–456. doi:10.1080/1525000902996532

Hollig, G. (2005). Supercoder: A program for coding preferential looking (Version 1.5). [Computer Software]. West Lafayette, IN: Purdue University.

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and segment processing in Mandarin idiom. Psychophysiology, 49, 1179–1190. doi:10.1111/j.1469-8986.2012.01406.x

Liu, L., & Kager, R. (2017). Enhanced music sensitivity in 9-month-old bilingual infants. Cognitive Processing, 18, 55–65. doi:10.1016/j.cogpro.2016.07-080-7

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. Behavior Research Methods, 44, 1042–1051. doi:10.3758/s13428-012-0203-3

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. Language and Speech, 47, 109–138. doi:10.1177/00238309040470020101

Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. Language and Cognitive Processes, 22, 566-594. doi:10.1080/0169096060989600

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. Language Learning and Development, 7, 209–225. doi:10.1080/15475441.2011.579839

Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. Proceedings of the National Academy of Sciences, U.S.A., 112, 14563–14568. doi:10.1073/pnas.1515087112

Ma, W., Zhou, P., Crain, S., & Gao, L. (2017a). Lexical tones and word learning in Mandarin-speaking children at three years of age. Journal of Electronic Science and Technology, 15, 25–32.

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017b). Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. Cognition, 159, 139–155. doi:10.1016/j.cognition.2016.11.011

Malins, J., & Joanisse, M. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. Journal of Memory and Language, 62, 407–420. doi:10.1016/j.jml.2010.02.004

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. Journal of Memory and Language, 57, 252–272. doi:10.1016/j.jml.2007.03.005

Mattock, K., & Burnham, D. (2006). Chinese and English infants’ tone perception: Evidence for perceptual reorganization. Infancy, 10, 241–265. doi:10.1177/s1521176506273024

Mattock, M., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. Cognition, 106, 1367–1381. doi:10.1016/j.cognition.2007.07.002

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between
consonants and vowels. Cognition, 98, 13–30. doi:10.1016/j.cognition.2004.10.005
Quam, C., & Swingley, D. (2010). Phonological knowledge guides two-year-olds’ and adults’ interpretation of salient pitch contours in word learning. Journal of Memory and Language, 62, 135–150. doi:10.1016/j.jml.2009.09.003
Repp, B., & Lin, H. (1990). Integration of segmental and tonal information in speech perception: Across-linguistic study. Journal of Phonetics, 18, 481–495.
Schirmer, A., Tong, S., Penney, T., Gunter, T., & Chen, H. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. Journal of Cognitive Neuroscience, 17, 1–12. doi:10.1162/0898929052880057
Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. Cognition, 124, 128–142. doi:10.1016/j.cognition.2012.05.008
Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. Cognition, 142, 1–11. doi:10.1016/j.cognition.2015.05.010
Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. Developmental Science, 17, 94–109. doi:10.1111/desc.12097
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. Cognition, 76, 147–166. doi:10.1016/S0010-0277(00)00081-0
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science, 13, 480–484. doi:10.1111/1467-9280.00485
Taft, M., & Chen, H. (1992). Judging homophony in Chinese: The influence of tones. In H. Chen & O. J. L. Tseng (Eds.), Language processing in Chinese (pp. 151–172). Oxford, England: North-Holland.
Tardif, T., Fletcher, P., Zhang, Z. X., Liang, W. L., & Zuo, Q. H. (2008). The Chinese communicative development inventory (Putonghua and Cantonese versions): Manual, forms, and norms. Beijing: Peking University Medical Press.
Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. Language and Cognitive Processes, 23, 689–708. doi:10.1080/01690960701728261
van Heuven, V. J., & Haan, J. (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven & N. Warner (Eds.), Papers in laboratory phonology VII (pp. 61–86). Berlin: Mouton de Gruyter.
Werker, J. F., Pegg, J. E., & McLeod, P. (1994). A cross-language comparison of infant preference for infant-directed speech: English and Cantonese. Infant Behavior and Development, 17, 321–331. doi:10.1016/0163-6383(94)90012-4
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. Journal of Memory and Language, 59, 114–132. doi:10.1016/j.jml.2008.03.001
Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. Language and Speech, 59, 59–82. doi:10.1177/0023830915578000
Ye, Y., & Connine, C. (1999). Processing spoken Chinese: The role of tone information. Language and Cognitive Processes, 14, 609–630. doi:10.1080/02699930260260202
Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. Journal of Memory and Language, 68, 123–139. doi:10.1016/j.jml.2012.09.004
Yip, M. J. W. (2002). Tone. Cambridge and New York: Cambridge University Press.
Yuan, J. (2011). Perception of intonation in Mandarin Chinese. The Journal of the Acoustical Society of America, 130, 4063–4069. doi:10.1121/1.3651818
