

Lexical tone variation and spoken word recognition in preschool children: effects of perceptual salience*

LEHER SINGH, ALOYSIA TAN AND
THILANGA D. WEWALAARACHCHI

Department of Psychology, National University of Singapore

*(Received 9 December 2015 – Revised 21 March 2016 – Accepted 19 May 2016 –
First published online 5 July 2016)*

ABSTRACT

Children undergo gradual progression in their ability to differentiate correct and incorrect pronunciations of words, a process that is crucial to establishing a native vocabulary. For the most part, the development of mature phonological representations has been researched by investigating children's sensitivity to consonant and vowel variation, with a much lesser focus on lexical tones. The current study investigates sensitivity to lexical tones in word recognition with specific attention to role of perceptual salience. Chinese-speaking preschoolers were presented with familiar words that were correctly pronounced, substituted for a subtle tone variant (Tones 2 and 3), or substituted for a salient tone variant (Tones 1 and 4). Results demonstrated that subtle tone variants were mistakenly perceived as correct pronunciations and only salient tone variants were recognized as mispronunciations. Findings suggest that tone integration follows a more complex developmental course than previously concluded.

INTRODUCTION

In order to recognize spoken words, children must harbor a very precise set of expectations about the relationship between sound and meaning. Formulating this set of expectations is complicated by the inordinate variability intrinsic to human speech. Learners must learn from experience how to assign appropriate relevance to the full range of variation in speech, gradually converging upon a phonological system that is closely aligned with the structure of the native lexicon. Phonological specification within the developing lexicon has been extensively studied in Romance

[*] We are grateful to an Academic Research Fund grant (MOE Tier 1 FY2013-FRC2-009) to support the research reported in this paper. Address for correspondence: Leher Singh, Dept. of Psychology, National University of Singapore, 9 Arts Link, Singapore 117570. e-mail: leher.singh.nus@gmail.com

and Germanic languages, such as English, Spanish, and French (e.g. Mani & Plunkett, 2007; Nazzi, 2005; Swingley & Aslin, 2000, 2002; White & Morgan, 2008). In contrast, such investigations remain a relative rarity in tone languages, yet tone languages comprise the linguistic majority. They are widely spoken around the world and are more frequent than non-tone (or intonation) languages (Fromkin, 1978; Yip, 2002). Moreover, over half of the world's population speak a native tone language (Fromkin, 1978), yet empirical studies of language acquisition focus predominantly on native learners of intonation languages. A by-product of a disproportionately weighty focus on languages such as English for understanding of the phonological lexicon is that effects of vowels and consonants on word recognition have been widely researched. In contrast, our understanding of the consequences of tone variation on word recognition remains limited. The purpose of the current study is to investigate the specificity of phonological representations for tones by exploring children's responses to different types of lexical tone variation in spoken word recognition.

Tone languages are defined by tripartite phonological systems, which draw distinctions in word meaning by varying three levels of the phonological code: consonants, vowels, and lexical tone. The last source of variation, lexical tone, entails syllable-level shifts in fundamental frequency (or in its perceptual correspondent, vocal pitch), amplitude, and duration (Blicher, Diehl & Cohen, 1990; Edmondson & Esling, 2006; Leather, 1983; Liu & Samuel, 2004; Whalen & Xu, 1992; Wong & Diehl, 2003). The most widely spoken tone language, Mandarin Chinese, has four lexical tones: Tone 1 (high level tone), Tone 2 (rising tone), Tone 3 (dipping tone), and Tone 4 (falling tone) (see Figure 1 for a depiction of Mandarin Chinese tones). Each tone communicates word meaning in conjunction with vowels and consonants. For example, the word *ma* assumes different meanings based on the tone in which it is produced. *Ma* means 'mother' when spoken in Tone 1, 'hemp' when spoken in Tone 2, 'horse' when spoken in Tone 3, and 'to scold' when said in Tone 4. Learners of a tone language therefore have to simultaneously track meaningful variation in vowels, consonants, and tones to arrive at the phonological determinants of meaning.

Traditionally, in experimental research, children's abilities to track phonological determinants of meaning were studied by investigating the accuracy and efficiency with which they recognized spoken words. More specifically, measuring children's sensitivity to words that deviate from their correct form via a phonological substitution (e.g. fixation to an image of a cup upon hearing *tʌp* versus *cʌp*) have provided us with a powerful means with which to scrutinize the early phonological lexicon and the level of specificity therein. Studies in this area, notably mispronunciation studies, have contributed significantly to our understanding of the degree of phonological definition associated with the nascent lexicon. These investigations have also

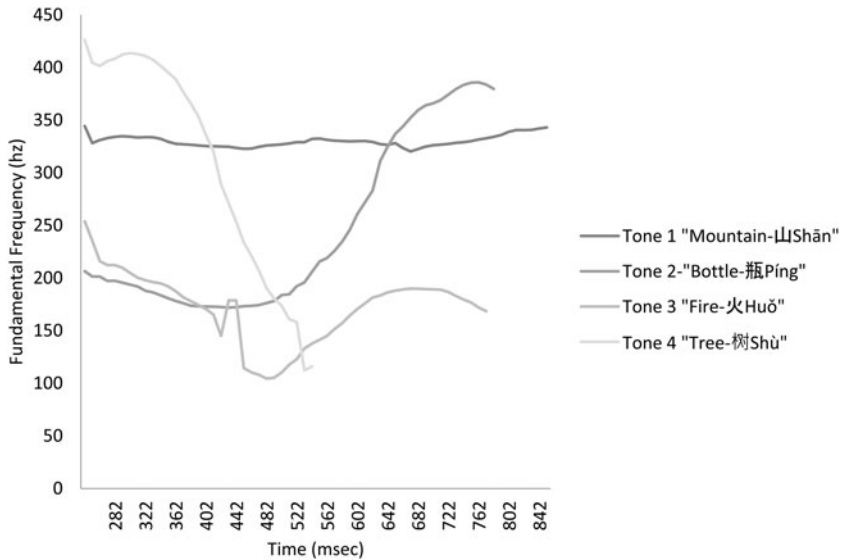


Fig. 1. Pitch contours for Mandarin tones.

revealed some of the constraints on children's sensitivity to mispronunciations, reporting moderating effects of the particular task employed (e.g. preferential looking versus categorization tasks), the specific contrast used, the extent of deviation from the correct pronunciation, and of the age at which children are tested (e.g. Curtin, Fennell & Escudero, 2009; Havy, Bertocini & Nazzi, 2011; Mani, Coleman & Plunkett, 2008; Mani & Plunkett, 2008, 2011; Singh, Goh & Wewalaarachchi, 2015; Swingley & Aslin, 2000, 2002; Van der Feest, 2007; White & Morgan, 2008).

For the most part, mispronunciation studies have focused on manipulating vowels and consonants within familiar words and comparing children's responses to correct and mispronounced words. A much more marginal emphasis has been placed on researching how learners of tone languages represent lexical tones in their early lexicon. Thus far, experimental studies on sensitivity to lexical tone in native learners of tone languages have focused on whether infants can discriminate lexical tones via habituation paradigms, rather than on the strength with which tone is represented in developing lexical representations. Discrimination studies have demonstrated that the ability to distinguish lexical tones is observed in tone language learning infants within the first four to six months of life (Mattock & Burnham, 2006; Yeung, Chen & Werker, 2013), suggesting that sensitivity to tone contrasts is solidified in the tone learner prior to vowels and consonants (Yeung *et al.*, 2013).

A question to arise from the conclusion that tone sensitivity emerges early in infancy is whether this high degree of sensitivity applies across the native tone inventory or whether it is specific to particular tones. Tsao (2008) investigated Mandarin-learning infants' abilities to discriminate different pairs of lexical tones at 10–12 months of age, revealing that infants' abilities to discriminate lexical tones was mediated by the particular tone contrast involved: salient Mandarin tone contrasts such as Tones 1 (high level) and 3 (dipping) tones were discriminated more accurately than very similar tones such as Tone 2 (rising) and Tone 3 (dipping). Similar asynchronies in tone discrimination have also been observed in early childhood (Eliot, 1991; Hao, 2012; Kiriloff, 1969; Wong, Schwartz & Jenkins, 2005; Wong 2012). This suggests that individual tones may be represented with unequal strength in infants' learning systems. Each of these prior studies has revealed different degrees of sensitivity to tone variation based on properties of individual tone pairs. Specifically, as the similarity between tones increases, sensitivity to the tone contrast has been shown to increase incrementally (e.g. Tsao, 2008; Wong 2005, 2012), a phenomenon also documented in studies investigating sensitivity to phonetic segments such as consonants (White & Morgan, 2008).

Prior studies with phonetic segments, such as consonants, invite the strong possibility that patterns observed in auditory discrimination tasks do not always predict behavior in word learning tasks. Specifically, successfully discriminated sounds in infancy are not always discretely bound to lexical representations in early childhood (e.g. Stager & Werker, 1997; Werker, Fennell, Corcoran & Stager, 2002). It is well attested that infants may be highly sensitive to particular sources of sound variation in a non-lexical context, but when they are faced with the added, simultaneous demands of linking the same sources of variation to meaning, they can exhibit a lesser awareness of phonological contrast (see Curtin, Byers-Heinlein & Werker, 2011, for a discussion of these issues).

There has been one prior study investigating tone sensitivity in word recognition in a sample of tone language learners. In a study by Singh *et al.* (2015), two groups of Mandarin-learning children (toddlers and preschoolers) were presented with familiar words that were either correctly pronounced or altered via a consonant, vowel, or tone mispronunciation. Each vowel and consonant mispronunciation entailed a single feature substitution and tone mispronunciations entailed tone substitutions between Tones 1, 2, and 4. Children were tested at two age groups (2;5 to 3;5 years and 4 to 5 years). At both age groups, children showed similar degrees of sensitivity to vowel and consonant substitutions and a distinct pattern of results for tone substitutions. This strongly suggests that it may not be viable to generalize from the wealth of knowledge on consonant and vowel representation to lexical tones, which may follow a unique course of

acquisition. In particular, toddlers demonstrated a very strong sensitivity to tone substitutions as compared with vowels and consonants, leading the authors to conclude that at this age children are very sensitive to tones as a source of lexical contrast. This complements findings from infant auditory discrimination, revealing that native sensitivity to lexical tones emerges precociously, relative to vowels and consonants (see Yeung *et al.*, 2013). Older children, however, demonstrated relatively strong mispronunciation effects for vowels and consonants as compared with the toddler sample. However, they demonstrated relatively weak mispronunciation effects for tones, although tone mispronunciations were reliably detected. The authors attributed this attenuation in tone discrimination to a growing appreciation for the multiplex of functions served by pitch in language, such as conveyance of emotional prosody, emphatic stress, questions versus statements. The functional differentiation of pitch may be a late developing ability that emerges in the preschool years, as suggested by prior studies (Quam & Swingley, 2012; Singh & Chee, 2016). The early attentiveness to lexical pitch reported in the toddler sample, however, is of direct relevance to the current study. In particular, conclusions ventured by Singh *et al.* (2015), that tone is preferentially encoded in early word representations in toddlers, were based on an aggregate response to tone mispronunciations, collapsing across different tone contrasts. Each of the tone mispronunciations entailed relatively salient contrasts. Moreover, the most difficult tone contrast to discriminate – Tones 2 and 3 – was not included in this study. Tones 2 and 3 present an interesting pair to investigate as they are acoustically similar and they are also related via phonological rules. Specifically, the Tone 3 sandhi rule prescribes that when two syllables marked with Tone 3 appear in direct succession, the first syllable is substituted by Tone 2. Consequently, under particular contexts, Tone 2 can be considered an allotone of Tone 3 (Chen, 2000). As a result of acoustic similarity and joint involvement in tone sandhi, Tones 2 and 3 present an interesting point of a comparison to more distal tone pairs. The current study investigates effects of variation caused by subtle tone changes (Tones 2 and 3) and more salient tone changes (Tones 1 and 4). The current study therefore aims to scrutinize the strong sensitivity reported for tones in prior studies (e.g. Singh *et al.*, 2015; Yeung *et al.*, 2013), by investigating the influence of different tone pairings on children's sensitivity to mispronunciations of tone.

In the present study, Mandarin-learning preschoolers were presented with a series of familiar words. Words were either correctly pronounced or mispronounced via tone substitutions. Tones were substituted by interchanging the target tone with a dissimilar tone (alternation between Tones 1 and 4) or with a similar tone (alternation between Tones 2 and 3). Substitutions were made in both directions (e.g. a word marked by

Tone 1 was mispronounced using Tone 4, and in other trials a word marked by Tone 4 was mispronounced using Tone 1). Children's responses to similar versus dissimilar tone substitutions were compared to their responses to correct pronunciations.

METHOD

Participants

Twenty-five three-year-old native speakers of Mandarin Chinese (12 boys) participated in the current study (mean age: 38 months 11 days, age range: 35 months 5 days to 40 months 1 day). Two additional participants were excluded from the final sample for inattention (failure to complete the experiment). All participants were typically developing children, performing at grade level with no known developmental disabilities or delays.

Stimuli

Auditory stimuli. Twenty-four imageable monosyllabic common nouns (i.e. 6 tokens from each lexical tone category: high, rising, dipping, and falling) were selected as stimuli. A female native speaker of Mandarin Chinese recorded all the tokens in a sound-attenuated recording booth in an infant-directed register. All target words were presented in sentence-final position with the carrier phrase: “你看, [target]!” (‘Look! [target]!’). The total pool of twenty-four nouns were distributed into three versions of the experiment (with 8 words per condition). Within each version, each of the eight words was either correctly pronounced (8 trials) or mispronounced (8 trials). The difference between experimental conditions was only the stimulus set used; three conditions were created to ensure a reasonably large set of stimuli across all participants and to ensure that conclusions and generalizations were not drawn based on a relatively small set of eight lexical items (see Table 1 for a list of all stimuli and versions).

Visual stimuli. Visual stimuli consisted of photographed images of targets and distractors. All visual stimuli were presented against a white background. Each image was scaled to 200 × 200 pixels. Areas of interest (AoIs) were set at 320 × 320 pixels to ensure that each visual stimulus was contained within the AoIs. Images were horizontally aligned. The side of presentation of the target was counterbalanced within subjects. The target image was always a familiar object and the distractor image was always an unfamiliar object. This decision was made on the grounds that target selection in the face of familiar target and distractor pairings can be influenced by the phonological properties of the target as well as the properties of the distractor. In other words, target selection could arise

TABLE 1. *Stimuli for each experimental condition (target objects are in boldface)*

| Practice trials (identical for each version) | | Target | Distractor | Auditory label |
|--|--|-------------|-------------|----------------|
| | | Pear | Biscuit | Lí |
| | | Clock | Cake | Zhōng |

| Test trials (Version 1) | Target | Distractor | Original tone | Correct pronunciation or substituted tone | |
|----------------------------|-----------------|--------------|------------------|--|-----------------------|
| | Mountain | Gourd | Shān | 1 | Correct pronunciation |
| | T-shirt | Mortar | Yī | 1 | Correct pronunciation |
| | Goat | Castanets | Yáng | 2 | Correct pronunciation |
| | Bottle | Kettle | Píng | 2 | Correct pronunciation |
| | Horse | Otter | Mǎ | 3 | Correct pronunciation |
| | Ear | Coffee bean | ěr | 3 | Correct pronunciation |
| | Tree | Net | Shù | 4 | Correct pronunciation |
| | Shop | Saron | Diàn | 4 | Correct pronunciation |
| | Mountain | Chess piece | Shàn | 1 | 4 |
| | Tree | Bead | Shū | 4 | 1 |
| | Goat | Door stopper | Yǎng | 2 | 3 |
| | Horse | Stingray | Má | 3 | 2 |
| | Bottle | Wheat | Píng | 2 | 3 |
| | Shop | Yarn | Diān | 4 | 1 |
| | T-shirt | Drum | Yì | 1 | 4 |
| | Ear | Oven glove | ěr | 3 | 2 |

| Test trials (Version 2) | | Target | Distractor | Original tone | Correct pronunciation or substituted tone |
|----------------------------|--------------------------------|--------------|------------|------------------|--|
| | Bun | Mortar | Bāo | 1 | Correct pronunciation |
| | Ice | Gourd | Bǐng | 1 | Correct pronunciation |
| | Cloud | Castanets | Yún | 2 | Correct pronunciation |
| | Snake | Kettle | Shé | 2 | Correct pronunciation |
| | Fire | Otter | Huó | 3 | Correct pronunciation |
| | Hand | Coffee bean | Shǒu | 3 | Correct pronunciation |
| | Moon | Saron | Yuè | 4 | Correct pronunciation |
| | Kai lan (vegetable) | Net | Cài | 4 | Correct pronunciation |
| | Ice | Chess piece | Bǐng | 1 | 4 |
| | Bun | Drum | Bào | 4 | 1 |
| | Cloud | Door stopper | Yún | 2 | 3 |
| | Fire | Stingray | Huó | 3 | 2 |
| | Hand | Oven glove | Shǒu | 3 | 2 |
| | Moon | Yarn | Yuē | 4 | 1 |
| | Snake | Wheat | Shě | 2 | 3 |
| | Kai lan (vegetable) | Bead | Cāi | 4 | 1 |

Table 1 (cont.)

| Test trials (Version 2) | | | | |
|----------------------------|--------------|-------|---|-----------------------|
| Star | Gourd | Xīng | 1 | Correct pronunciation |
| Porridge | Mortar | Zhōu | 1 | Correct pronunciation |
| Monkey | Castanets | Hóu | 2 | Correct pronunciation |
| Insect | Kettle | Chǒng | 2 | Correct pronunciation |
| Ocean | Coffee bean | Hǎi | 3 | Correct pronunciation |
| Eye | Otter | Yán | 3 | Correct pronunciation |
| Road | Net | Lù | 4 | Correct pronunciation |
| Chicken | Saron | Ròu | 4 | Correct pronunciation |
| Star | Chess piece | Xìng | 1 | 4 |
| Porridge | Drum | Zhòu | 1 | 4 |
| Ocean | Oven glove | Hái | 3 | 2 |
| Monkey | Door stopper | Hǒu | 2 | 3 |
| Eye | Stingray | Yán | 3 | 2 |
| Road | Bead | Lū | 4 | 1 |
| Chicken | Yarn | Rōu | 4 | 1 |
| Insect | Wheat | Chǒng | 2 | 3 |

because there is a match between the auditory input and the label stored for the target object, or because there is a mismatch between the auditory input and the label stored for the distractor object. It is impossible to determine which of these processes prevails for a participant within any given trial (see White & Morgan, 2008, for a discussion of this issue). It should be noted, however, that Singh *et al.* (2015) demonstrated no effects of distractor familiarity on tone mispronunciation effects in toddlers.

Stimulus validation. To ensure that lexical identity of the tokens was conveyed as intended, acoustic analyses were conducted on all recorded stimuli to obtain pitch characteristics of the target words. For each token, mean fundamental frequency (F_0), fundamental frequency at onset and offset, and fundamental frequency range were calculated and averaged across lexical tone categories. For Tone 1 (high tone), there was minimal pitch variation from syllable onset to syllable offset (F_0 range: 333 Hz to 359 Hz) and a high overall mean pitch (337 Hz). Tone 2 (rising tone) showed the largest increase in pitch from syllable onset to syllable offset (220 Hz to 378 Hz; mean F_0 : 230 Hz). Tone 3 (dipping tone) showed an intermediate onset pitch (223 Hz) followed by an inflection point (112 Hz) and ended with a terminal rise to 228 Hz. Tone 4 (falling tone) began with the highest onset pitch (416 Hz) and ended with a low terminal pitch (173 Hz). All tokens were consistent with expected pitch contours for Mandarin tones. Mean pitch, minimum pitch, onset and offset pitch, and pitch range did not vary based on whether the tone-bearing syllable was a correct pronunciation or a tone substitution (all p values > .7). Twelve

native adult speakers of Mandarin Chinese were asked to categorize stimuli into lexical tone categories. All tokens were identified with high accuracy: Tone 1 (99%), Tone 2 (93%), Tone 3 (91%), and Tone 4 (97%).

Apparatus and procedure

The Preferential Looking Paradigm was employed in the current study as realized in prior studies to investigate sensitivity to mispronunciations of familiar words (e.g. Mani & Plunkett, 2010; Singh *et al.*, 2015; White & Morgan, 2008). The Tobii 60XL eye-tracking system (Version 3.2.1) was used. All participants were seated 70 cm away from a 24-inch Tobii 60XL monitor, which was placed comfortably at eye level. Auditory stimuli were presented at a conversation level via left–right speakers embedded within the monitor. An experimenter manually coded 25% of the data obtained from Tobii. Inter-coder reliability was .94.

The experiment began with two training trials that served to familiarize children to the task followed by sixteen test trials in a randomized order. Each test trial was split into two phases of equal duration – the pre-naming and the post-naming phase, with the entire trial lasting 5000 ms. The visual display consisted of one familiar target object and a novel distracter object, both of which stayed on screen for the entire duration of both phases. In each trial, children heard the directive “你看!” (‘Look!’) during the pre-naming phase. For each trial, the auditory stimulus was synchronized to begin 2500 milliseconds after the start of each trial, initiating the post-naming phase.

Children were presented with three types of test trials: words that were correctly pronounced ($n = 8$), words that underwent acoustically distinct mispronunciations ($n = 4$; two trials involving substitutions from Tone 1 to Tone 4, and two trials involving substitutions from Tone 4 to Tone 1), and words that underwent acoustically subtle mispronunciations ($n = 4$; two trials involving Tone 2 to Tone 3 substitutions, and two trials involving Tone 3 to Tone 2 substitutions). Each participant heard the same word correctly pronounced and mispronounced. Half of the mispronunciations were subtle mispronunciations (from Tone 2 to Tone 3 and vice versa), and half were salient mispronunciations (from Tone 1 to Tone 4 and vice versa).

RESULTS

As in prior mispronunciation studies, the dependent measure was the differential in fixation to the target object prior to and following presentation of its label. Evidence of target recognition is typically inferred from the presence of a naming effect (Bailey & Plunkett, 2002; Meints, Plunkett & Harris, 1999; Schafer & Plunkett, 1998; Swingley & Aslin, 2000). Naming effects refer to a significant elevation in fixation to the

target object following naming. Naming effects are computed by the following formula: Proportion of Total Looking to the Target (PTL) during the post-naming phase minus Proportion of Total Looking to the Target during the pre-naming phase. The purpose of computing naming effects from both pre- and post-naming values is to mitigate effects of stimulus characteristics that may elicit preferential fixation independent of labeling. A significant positive naming effect created by an increase in PTL between the pre- and post-naming phases is typically recruited as evidence for children having associated the verbal label with the visual target within a trial. In contrast, a naming effect that does not deviate significantly from zero (i.e. no significant difference in PTL values between the pre- and post-naming phases) is presumed to indicate ambiguity with regards to the referent for the verbal label.

Naming effects were computed for correct pronunciations and for mispronunciations. As naming effects did not differ based on the direction of substitution (Tone 1 substituted for Tone 4 or vice versa; Tone 2 substituted for Tone 3 or vice versa; $p > .7$), they were collapsed across direction of substitution and computed for Tone 1–4 mispronunciations, Tone 2–3 mispronunciations, and correct pronunciations. Likewise, there were no differences in baseline fixation to targets during the pre-naming phase across trial types ($p > .7$). Data from 3% of trials were excluded due to failure to attend to the screen at all during individual test trials.

Naming effects for each trial type are displayed in Figure 2a. As depicted in Figure 2a, there was an elevation in fixation to the target (i.e. a naming effect) for correct pronunciations and for Tone 2–3 mispronunciations. There was no elevation in fixation to the target when Tones 1 and 4 were substituted. A series of one-sample *t*-tests confirmed that naming effects departed significantly from zero for correct pronunciations ($t(24) = 3.18$, $p = .004$; Cohen's $d = .85$). Likewise, naming effects departed significantly from zero for mispronunciations involving Tones 2 and 3 ($t(24) = 2.81$, $p = .01$; Cohen's $d = .79$). There was no significant change in naming effects in comparison to zero for substitutions involving Tone 1 and 4. All findings remained significant following a Bonferroni correction for multiple comparisons. This demonstrates that children preferentially fixated target objects even when they were mislabeled on account of a Tone 2/3 substitution. Naming effects were comparable when Tone 2 was substituted for Tone 3 and when Tone 3 was substituted for Tone 2. Naming effects were also comparable when Tone 1 was substituted for Tone 4 and vice versa (see Figure 2b). This was statistically confirmed by no effect of direction of substitution on naming effects for either type of mispronunciation ($p < .7$).

A repeated-measures analysis of variance was conducted to determine whether the strength of naming effects varied by trial type (correct

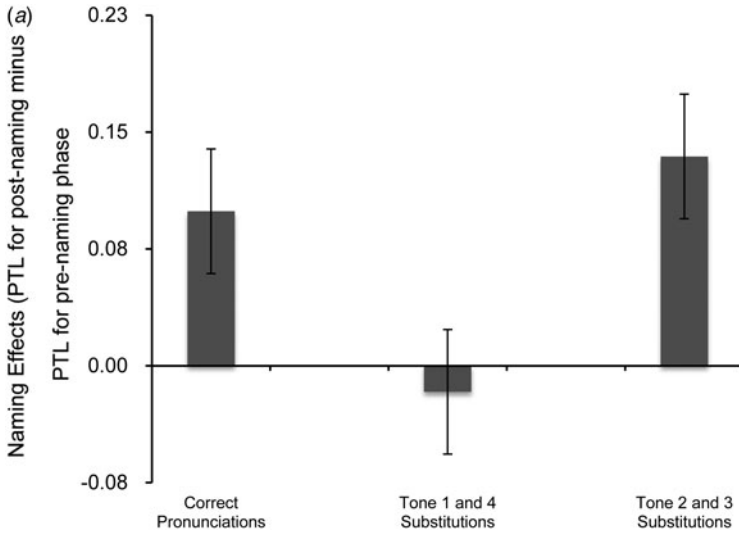


Fig. 2a. Naming effects for correct pronunciations, Tone 1–4 substitutions and Tone 2–3 substitutions (error bars: SEM).

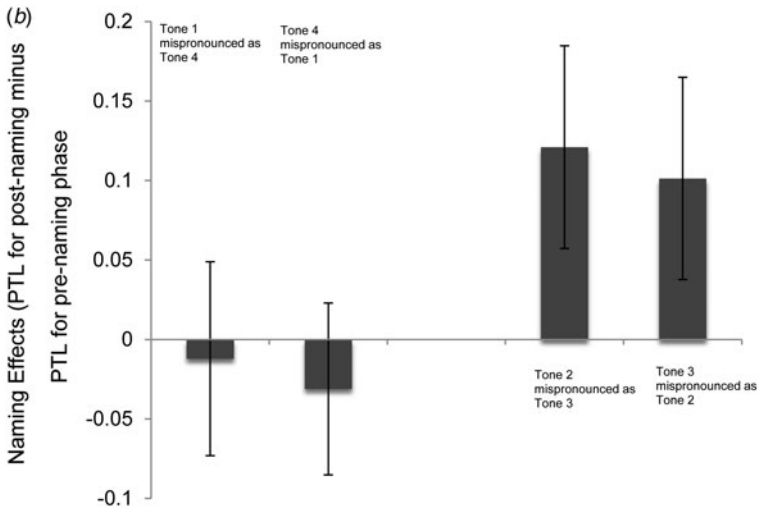


Fig. 2b. Naming effects for Tone 1–4 and Tone 2–3 substitutions (error bars: SEM).

pronunciations, Tone 1 and 4 mispronunciations, Tone 2 and 3 mispronunciations). Results revealed a main effect of trial type ($F(2,42) = 3.98$, $p = .04$, partial $\eta^2 = .24$). Post-hoc pairwise comparisons were

computed using Tukey's HSD test with a significance criterion of $p < .05$. Results revealed a significantly higher naming effect for correct pronunciations in comparison to Tone 1–4 substitutions (0.09 vs. -0.02), a significantly higher naming effect for Tone 2–3 substitutions in comparison to Tone 1–4 substitutions (0.13 vs. -0.02), but no significant difference in naming effects for Tone 2–3 substitutions and correct pronunciations (0.13 vs. 0.09). As such, it appears that a subtle tone contrast (Tone 2 and 3 substitutions) was interpreted akin to correct pronunciations. Moreover, mispronunciation effects were only evident when Tones 1 and 4 were substituted.

A second set of analyses was aimed at disaggregating naming effects to examine the temporal dynamics of word recognition. Prior studies have revealed that a more nuanced profile of lexical selection can be established by looking at the timecourse of word recognition (e.g. Fernald, McRoberts & Swingley, 2001). In particular, for the subset of trials on which infants are fixated on the distractor object at the start of the post-naming phase, the proportion of shifts to the target for each type of pronunciation can be used to determine the temporal processing 'cost' of mispronunciations (Fernald *et al.*, 2001). To examine the timecourse of lexical selection, we plotted shifts to the target, frame-by-frame, for all distractor-initial trials by trial type (see Figure 3). Distractor-initial trials constituted 46% of trials. As depicted in Figure 3, there are apparent differences in shifts to the target for Tone 2–3 mispronunciations in comparison to correct pronunciations that are not revealed by PTL. These differences were statistically scrutinized by segmenting the entire test block into windows of analysis (epochs) of 100 milliseconds each, resulting in 25 contiguous epochs. The first 200 milliseconds was excluded from the window of analysis on account of the fact that re-fixation of eye-gaze from one location to another takes approximately 200 milliseconds (Purves, Augustine, & Fitzpatrick, 2001). For each epoch, proportion of shifts to the target was compared for correct pronunciations, Tone 2–3 mispronunciations, and separately for correct pronunciations and for Tone 1–4 mispronunciations. In a series of pairwise comparisons, results revealed a higher proportion of shifts to the target in correct pronunciation trials as compared with Tone 1–4 mispronunciations for each 100-millisecond epoch (200–300, 300–400, and so on to 2500 ms; $p < .000001$ for all epochs). All comparisons remained significant following a Bonferroni correction for multiple comparisons, adopting a significance criterion of .05. These findings align with the results of PTL demonstrating reduced target fixation in the face of salient mispronunciations.

A parallel set of analyses was conducted to compare the proportion of shifts for Tone 2–3 mispronunciations versus correct pronunciations. This analysis is of primary interest because it has the potential to reveal a

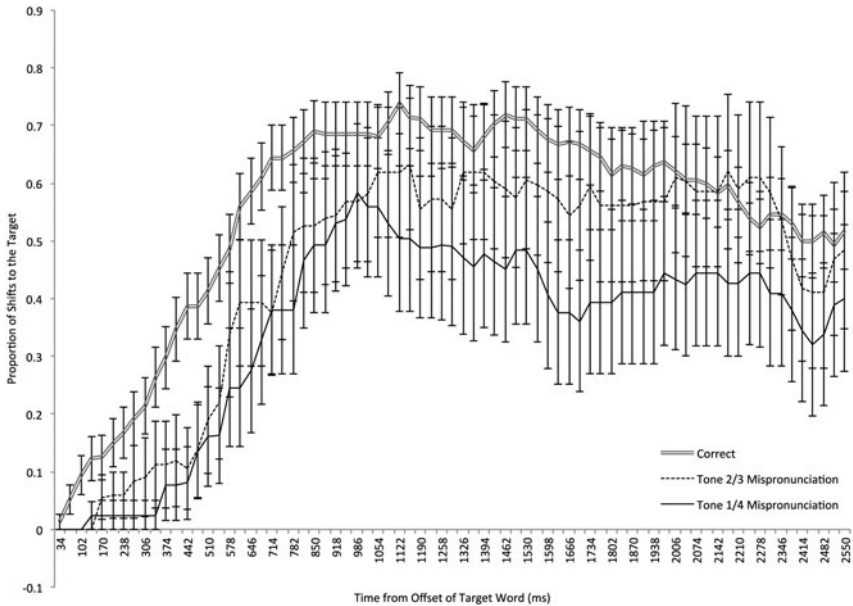


Fig. 3. Shifts to the target (distractor-initial trials) for correct pronunciations, Tone 1–4 substitutions and Tone 2–3 substitutions (error bars: SEM).

processing cost attached to a subtle mispronunciation, or whether subtle tone substitutions are truly processed akin to correct pronunciations. A series of *t*-tests drawing pairwise comparisons across 100-millisecond epochs were computed, subtracting the first 200 ms to allow for re-fixation. For a contiguous block of 16 epochs (200–300 to 1800–1900 ms), there was a significant increase in proportion of shifts to the target for correct pronunciations versus Tone 2–3 mispronunciations ($p < .00005$ for all epochs). After 1900 milliseconds, there was a convergence in proportion of fixation to the target for correct pronunciations and Tone 2–3 mispronunciations, and there were no significance differences in proportion of fixations to the target between 1900 and 2500 ms. All significant comparisons remained so following a Bonferroni correction for multiple comparisons, adopting a significance criterion of .05.

In combination, our results invite a slightly different account of tone sensitivity to subtle tone variation based on whether we rely on naming effects or on the timecourse of word recognition. Naming effects suggest that children ‘false-alarmed’ to Tone 2–3 substitutions, fixating the target object when Tone 2 was substituted for Tone 3 and vice versa. In contrast, a more pronounced sensitivity was observed for the more salient

contrast involving Tone 1 and 4 substitutions. However, an analysis of the timecourse of lexical selection revealed a more nuanced picture with regards to sensitivity to Tone 2–3 substitutions: for about 70% of the post-naming block, participants were less likely to fixate the target object when Tones 2 and 3 were substituted than when the object was correctly specified for tone. This suggests that there is some degree of sensitivity to subtle tone contrasts, and that subtle variations are not processed identically to correct pronunciations. However, salient tone contrasts appear to have been robustly rejected as possible labels for familiar target words, as revealed by naming effects and timecourse measures.

DISCUSSION

The purpose of the current study was to examine the phonological precision with which children represent lexical tones in spoken language processing. Native speakers of Mandarin Chinese were presented with a series of familiar words, some of which were correctly pronounced and some of which were mispronounced. Mispronounced words involved a substitution of either highly discriminable tones (Tones 1 and 4) or highly similar tones (Tones 2 and 3). Results demonstrated that children reliably recognized correctly pronounced forms. However, they also mis-identified substitutions of Tones 2 and 3 as correct labels for visual targets. In contrast, substitutions of Tones 1 and 4 were not identified as labels for visual targets. At first glance, these findings suggest that participants were categorically insensitive to the distinction between subtle tone contrasts in spoken word recognition. However, a more detailed analysis of the timecourse of lexical selection revealed that children were somewhat sensitive to Tone 2–3 substitutions, exemplified by a reduced proportion of shifts to the target for these trials relative to correct pronunciations for a substantial duration of the post-naming test block. By contrast, in comparison to correct pronunciation trials there was a persistent reduction in shifts to the target for Tone 1–4 substitutions throughout the entire post-naming test block.

Our overall pattern of results suggests that spoken language comprehension of tone-bearing units is heavily influenced by the perceived similarity of individual tones. In particular, the mis-identification of visual targets as potential referents for mispronounced forms of Tone 2 or Tone 3 points to a late-emerging sensitivity to subtle tone distinctions in spoken word recognition. In perceptual discrimination tasks, tone-learning infants at 10–12 months of age were able to distinguish this contrast, albeit less so than more salient contrasts (Tsao, 2008). The present set of findings suggests that infant tone discrimination abilities do not infiltrate the process of lexical selection and successfully discriminated contrasts are not necessarily integrated into the process of later word recognition.

In the present study, there were strong differences in children's sensitivity to correct and subtle tone contrasts (which were treated very similarly to one another) and salient tone contrasts, which were treated distinctly to correct/subtle mispronunciations. Results observed herein are somewhat consistent with prior studies on vowel variation on spoken word recognition in toddlers, demonstrating overlapping sensitivity to correct pronunciations and subtle mispronunciations and sharply deviating sensitivity to salient mispronunciations (e.g. Mani & Plunkett, 2011). As mentioned in the 'Introduction', Tones 2 and 3 are not only similar but are also subject to a process of phonological neutralization via tone sandhi. It is therefore possible that confusion between these tones arises from this neutralization. However, if this were the case, one would expect directional asymmetries in mispronunciation effects, as there is no situation that licenses a reversed substitution from Tone 2 to Tone 3. Specifically, mispronunciation effects would only be predicted when Tone 2 was substituted for Tone 3, and naming effects would be predicted when Tone 3 was substituted for Tone 2 as the latter substitution reflects the alternation associated with the Tone 3 sandhi rule. We found no evidence of directional effects on naming effects for Tone 2–3 substitutions, as shown in Figure 2b. Moreover, prior studies on spoken word recognition of sandhi forms suggest that children are not sensitive to Tone 3 sandhi rules until four to five years of age (Wewalaarachchi & Singh, 2014), pointing to the possibility that the conflation of these forms may be primarily due to their perceptual similarity and less so to their potential to be neutralized.

Previous investigations of tone representation across a broad range of discrimination and word recognition tasks would suggest that tone-exposed infants are quite sensitive to tone changes as infants and toddlers, and that tone is preferentially encoded relative to vowels and consonants (Singh & Foong, 2012; Singh *et al.*, 2015; Singh, Hui, Chan & Golinkoff, 2014; Tsao, 2008; Yeung *et al.*, 2013). The chief contribution of the current study is to modify the conclusion that tone information is preferentially available to native learners. In fact, sensitivity to subtle contrasts remains quite low even in preschoolers when children have typically amassed substantial vocabularies. These findings suggest that strong sensitivity to tone as a source of lexical contrast evinced in prior studies (Singh & Foong, 2012; Singh *et al.*, 2015) may be specific to highly contrastive tones.

Although tone integration appears to depend on the specific tone pairs involved, this raises the question of the types of cues that learners may profit from in learning more difficult contrasts. One possibility is that the availability of visual cues to tone may promote differentiation of these tones. Tone 3, although similar to Tone 2 in its pitch profile, is accompanied by distinctive facial movements in native speakers (a head and chin dip) which are not applied when native speakers produce Tone 2

(Chen & Massaro, 2008). These movements may help to differentiate similar tones in an interactive context. Research with adult native speakers of tone languages indicates a selective underutilization of visual cues when processing lexical tone information (i.e. performance is not augmented when comparing auditory-only and auditory-visual conditions) (e.g. Burnham, Cioccia & Stokes, 2001; Chen & Massaro, 2008). However, it is possible that young children may orient more closely to visual cues in the face of phonological ambiguity when mastering words and tones. They may utilize these cues to distinguish subtle tone distinctions, a phenomenon previously demonstrated in English-learning children (Jeger, Damian, Spence, Tye-Murray & Abdi, 2009). Future research could contrast effects of tone similarity in an auditory-only versus multimodal dynamic context to determine whether disambiguation in children is facilitated by multimodal cues. Second, the present set of findings are highly consistent with those observed in tone productions. Tones are sporadically contrasted in vocalizations (Hua & Dodd, 2000; Li & Thompson, 1977; So & Dodd, 1995). In particular, dissimilar tones are contrasted in early productions, whereas similar tones such as Tone 2 and 3 take several years to differentiate in vocal productions of tones (Wong, 2012, 2013; Wong *et al.*, 2005), even though children can discriminate auditory tokens of these tones in infancy (Tsao, 2008). The point in development at which clearly differentiated productions of Tones 2 and 3 reach adult-like targets remains unknown. However, this ability in production appears not to be mature even as late as four or five years of age (Wong, 2013). Future research could focus on longitudinal analyses of production and perception of similar tones to determine the extent to which tone perception and production may reinforce each other. Cross-lagged models applied to perception and production growth trajectories could elucidate feedback mechanisms available to children as they learn to differentiate subtle tones in a lexical context.

The purpose of the current study was to examine children's abilities to integrate lexical tones during spoken word recognition via a mispronunciation paradigm, adding to an emerging focus on tone language acquisition. The representation of tones in early childhood remains elusive, due to a predominant emphasis on sensitivity to consonants and vowels in prior research. In the current study, we observed that sensitivity to tone variation is clearly evident when tone contrasts are salient. When tone contrasts are more subtle, however, preschool children appear not to be sensitive to tone variation even for familiar words. In summary, the current findings point to strong effects of acoustic similarity on children's abilities to integrate lexical tones as determinants of word meaning. Findings suggest that although infants and toddlers appear sensitive to tone relatively early in development, as suggested by prior studies, this

sensitivity may be modified by the particular tone contrast involved, and may not generalize across native tone inventories. By implication, future studies should assess tone sensitivity in children across a diversity of tone contrasts, as responses to different contrasts can invite very different conclusions on the timing of lexical tone acquisition.

REFERENCES

- Bailey, T. M. & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development* 17(2), 1265–82.
- Blicher, D. L., Diehl, R. L. & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics* 18(1), 37–49.
- Burnham, D., Ciocca, V. & Stokes, S. (2001). Auditory-visual perception of lexical tone. Paper presented at EUROSPEECH 2001.
- Chen, M. Y. (2000). *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.
- Chen, T. H. & Massaro, D. W. (2008). Seeing pitch: visual information for lexical tones of Mandarin-Chinese. *Journal of the Acoustical Society of America* 123(4), 2356–66.
- Curtin, S., Byers-Heinlein, K. & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics* 39(4), 492–504.
- Curtin, S., Fennell, C. & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science* 12(5), 725–31.
- Edmondson, J. A. & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23(2), 157–91.
- Eliot, C. E. (1991). The relationship between the perception and production of Mandarin tones: an exploratory study. *University of Hawai'i Working Papers in ESL* 10(2), 177–204.
- Fernald, A., McRoberts, G. W. & Swingle, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn & B. Hoehle (eds), *Approaches to bootstrapping: phonological, syntactic and neurophysiological aspects of early language acquisition*, 97–124. Amsterdam/Philadelphia: John Benjamins.
- Fromkin, V. A. (ed.) (1978). *Tone: a linguistic survey*. New York, NY: Academic Press.
- Hao, Y. C. (2012). Second language acquisition of Mandarin tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40(2), 269–79.
- Havy, M., Bertocini, J. & Nazzi, T. (2011). Word learning and phonetic processing in preschool age children. *Journal of Experimental Child Psychology* 108, 25–43.
- Hua, Z. & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of child language* 27(1), 3–42.
- Jerger, S., Damian, M., Spence, M., Tye-Murray, N. & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: a new multimodal picture word task. *Journal of Experimental Child Psychology* 102, 40–59.
- Kirilloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica* 20(2–4), 63–7.
- Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11(4), 373–82.
- Li, C. N. & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* 4(2), 185–99.
- Liu, S. & Samuel, A. G. (2004). Perception of Mandarin lexical tones when Fo information is neutralized. *Language and Speech* 47(2), 109–38.
- Mani, N., Coleman, J. & Plunkett, K. (2008). Phonological specificity of vowel contrasts at 18-months. *Language and Speech* 51(1/2), 3–21.
- Mani, N. & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language* 57(2), 252–72.

- Mani, N. & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science* **11**(1), 53–9.
- Mani, N. & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological Science* **21**(7), 908–13.
- Mani, N. & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition* **121**(2), 196–206.
- Mattock, K. & Burnham, D. (2006). Chinese and English infants' tone perception: evidence for perceptual reorganization. *Infancy* **10**(3), 241–65.
- Meints, K., Plunkett, K. & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology* **35**(4), 1072–8.
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition* **98**(1), 13–30.
- Purves, D., Augustine, G. & Fitzpatrick, D. (2001). Types of eye movements and their functions. In *Neuroscience*, 2nd ed. Sunderland, MA: Sinauer.
- Quam, C. & Swingle, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development* **83**, 246–50.
- Schafer, G. & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development* **69**(2), 309–20.
- Singh, L. & Chee, M. (2016). Rise and fall: effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics* **55**, 109–18.
- Singh, L. & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition* **124**(2), 128–42.
- 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.
- 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**(1), 94–109.
- So, L. K. & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language* **22**(3), 473–95.
- Stager, C. L. & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* **388**(6640), 381–2.
- Swingle, D. & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* **76**(2), 147–66.
- Swingle, D. & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science* **13**(5), 480–4.
- Tsao, F. M. (2008). The effect of acoustically similarity of lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology* **50**(2), 111–24.
- Van der Feest, S. V. (2007). *Building a phonological lexicon: the acquisition of the Dutch voicing contrast in perception and production*. Prince Productions. Online <<http://suzanne.vanderfeest.nl/OutlineDiss.pdf>>.
- Werker, J. F., Fennell, C. T., Corcoran, K. M. & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* **3**(1), 1–30.
- Wewalaarachchi, D. & Singh, L. (2014). Influences of tone sandhi on word recognition in preschool children. Poster presented at the Fifteenth Annual Conference of the International Speech Communication Association.
- Whalen, D. H. & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* **49**(1), 25–47.
- White, K. S. & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language* **59**(1), 114–32.
- Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* **40**(1), 141–51.
- Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *Journal of the Acoustical Society of America* **133**(1), 434–43.

- Wong, P. C. & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research* **46**(2), 413–21.
- Wong, P., Schwartz, R. G. & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research* **48**(5), 1065–79.
- 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**(2), 123–39.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.